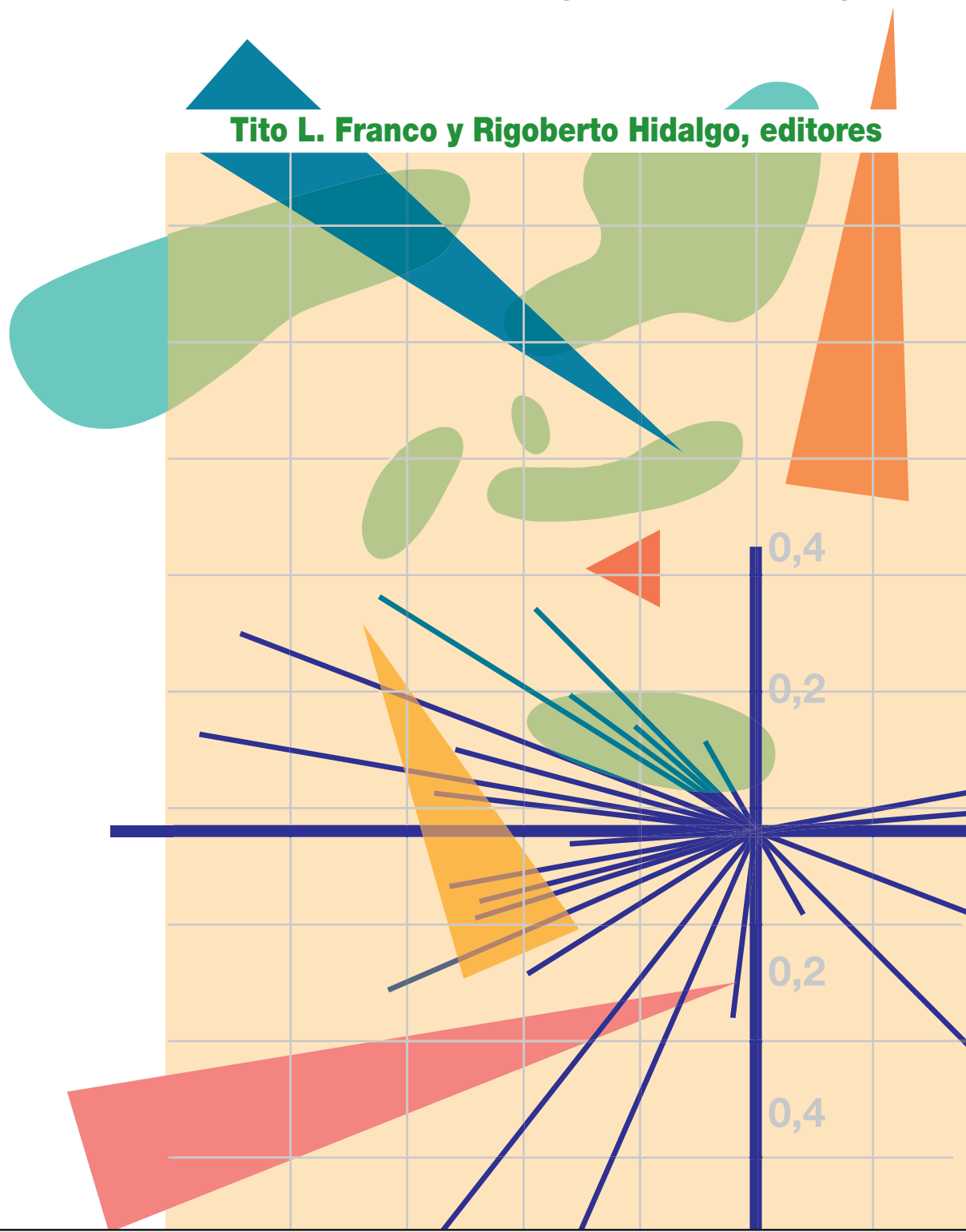




Análisis Estadístico de Datos

de Caracterización Morfológica de Recursos Fitogenéticos

Tito L. Franco y Rigoberto Hidalgo, editores



El Instituto Internacional de Recursos Fitogenéticos (IPGRI) es un organismo internacional autónomo de carácter científico que opera bajo los auspicios del Grupo Consultivo sobre Investigación Agrícola Internacional (GCAI). El mandato del IPGRI es promover la conservación y el uso de los recursos fitogenéticos para beneficio actual y futuro de la humanidad. El Instituto tiene su sede en Maccarese, cerca de Roma, Italia, y oficinas en más de 20 países del mundo. El IPGRI opera mediante tres programas: (1) el Programa de Recursos Fitogenéticos, (2) el Programa de Apoyo a las actividades en Recursos Fitogenéticos de los Centros del GCAI, y (3) la Red Internacional para el Mejoramiento del Banano y el Plátano (INIBAP).

El carácter de organismo internacional del IPGRI lo confiere la firma del Convenio de Creación del Instituto, a enero de 2002 ratificado por los gobiernos de los siguientes países: Argelia, Australia, Bélgica, Benin, Bolivia, Brasil, Burkina Faso, Camerún, Congo, Costa de Marfil, Costa Rica, Chile, China, Chipre, Dinamarca, Ecuador, Egipto, Eslovaquia, Grecia, Guinea, Hungría, India, Indonesia, Irán, Israel, Italia, Jordania, Kenia, Malasia, Mauritania, Marruecos, Noruega, Pakistán, Panamá, Perú, Polonia, Portugal, la República Checa, Rumania, Rusia, Senegal, Sudán, Suiza, Siria, Túnez, Turquía, Ucrania y Uganda.

Las designaciones geográficas empleadas en esta publicación, al igual que la presentación del material, no expresan en modo alguno opinión del IPGRI o del GCAI sobre el estatus legal de ningún país, territorio, ciudad o área, ni acerca de sus autoridades o de la delimitación de sus fronteras. Asimismo, las opiniones expresadas son las de los autores y no necesariamente reflejan los puntos de vista de estas organizaciones. La mención de alguna marca registrada se suministra con fines informativos únicamente, no de apoyo al producto.

Edición: Alberto Ramírez P. (Ing. Agr.)

Notación decimal: Cifras decimales con punto (ej., 0.33, 15.48); unidades de mil, sin signo hasta 9999 (ej., 2659); con coma desde 10,000 (ej., 18,950)

Cita:

Franco, T. L. e Hidalgo, R. (eds.). 2003. Análisis Estadístico de Datos de Caracterización Morfológica de Recursos Fitogenéticos. Boletín técnico no. 8, Instituto Internacional de Recursos Fitogenéticos (IPGRI), Cali, Colombia. 89 p.

ISBN 92-9043-543-7

IPGRI
Via dei Tre Denari 472/a
00057 Maccarese
Roma, Italia

© Instituto Internacional de Recursos Fitogenéticos, 2003

Introducción a la Serie Boletines Técnicos del IPGRI

Los Boletines Técnicos de la serie IPGRI son producidos por científicos que trabajan en el campo de los recursos fitogenéticos y están dirigidos a las personas encargadas del manejo de las colecciones de dichos recursos en las diferentes regiones del mundo. Cada título proporciona una guía que permite seleccionar y evaluar las técnicas y procedimientos de conservación, y conocer los requerimientos y opciones de experimentación adaptados a las condiciones de cada localidad y especies de interés.

El IPGRI promueve y apoya el intercambio de información sobre los resultados obtenidos por los investigadores en los laboratorios y bancos de germoplasma, a la vez que agradece las sugerencias sobre temas para próximos números de esta serie.

Presentación

En los bancos de germoplasma en los países de América del Sur, Centroamérica y el Caribe se encuentran aproximadamente 660,000 accesiones pertenecientes a un rango amplio de géneros y especies de plantas las cuales están sólo parcialmente documentadas. En los últimos años se han realizado esfuerzos importantes en el registro y almacenamientos de los datos de pasaporte y la caracterización de estas accesiones, así como en el intercambio de la información relacionada. No obstante, con algunas excepciones, existe un conocimiento limitado sobre cómo analizar estos datos, las técnicas y las herramientas estadísticas disponibles y los diferentes tipos de análisis posibles según las necesidades y los objetivos de los investigadores y botánicos.

Las publicaciones sobre los conceptos básicos de caracterización y los métodos de análisis estadístico de germoplasma por lo general contienen poca información y su empleo no resulta práctico para la mayoría de los usuarios. Aunque existen trabajos de tesis y artículos científicos en revistas especializadas, su forma es puntual y no destacan la importancia del análisis estadístico en la caracterización morfológica de los recursos genéticos, ni sirven de guía para que los investigadores seleccionen las técnicas apropiadas según sus necesidades.

El Instituto Internacional de Recursos Genéticos (IPGRI) con la colaboración de expertos desarrolló el presente boletín como una ayuda para investigadores y curadores sobre las alternativas posibles cuando se analizan conjuntos de datos de caracterización morfológica de germoplasma. Es importante, sin embargo, tener en cuenta que no se trata de un manual de estadística ni de manejo de herramientas o programas relacionados. La idea principal es proporcionar una síntesis de las técnicas actualmente disponibles para analizar datos de caracterización morfológica y un marco de referencia para la identificación de las técnicas más apropiadas en cada caso.

Para la preparación de este boletín se tuvieron en cuenta varios Estudios de Casos en los que se incluyen como objetivos de la caracterización: (1) la medición de la variabilidad genética de la colección de germoplasma; (2) el establecimiento de la representatividad de los ejemplares en la colección en relación con la variabilidad del germoplasma en una región o con la variabilidad total de la especie involucrada; y (3) la caracterización de la estructura genética. Otros objetivos, no utilizados en los estudios de casos en este boletín, son la identificación de duplicados en una colección y la detección de genes especiales o alelos particulares, bien sea solos o en combinaciones únicas, que se expresan en caracteres morfológicos en diferentes estados.

En este boletín se hace énfasis en el análisis estadístico de los datos de caracterización morfológica como el primer paso para estudiar la variabilidad genética y se presentan, en forma general, las diferentes técnicas estadísticas que facilitan un mejor entendimiento de dicha variabilidad. Sin embargo, los editores quieren llamar la atención del investigador indicando que en la temática no se incluyen técnicas estadísticas relacionadas con la biología molecular la cual ha agregado nuevos conceptos y herramientas para un mejor entendimiento de la variabilidad y donde, igualmente, el análisis de datos juega un papel importante.

Aunque los actuales avances derivados de la biología molecular han proporcionado nuevas técnicas para la caracterización de germoplasma, la aplicación de éstas todavía

no se ha generalizado en la mayoría de los bancos de germoplasma de la región. Así mismo, los desarrollos en los sistemas de información geográfica (SIG) están facilitando la identificación de las condiciones predominantes en los sitios de origen de las especies vegetales. En consecuencia, gracias a estas herramientas modernas la caracterización está tomando una nueva orientación hacia el análisis integrado que incluye varias disciplinas para una mejor comprensión de la variabilidad genética de las colecciones de germoplasma.

Se espera que las técnicas aquí descritas sean útiles para los investigadores en recursos genéticos y curadores de colecciones de germoplasma que trabajan en forma exclusiva con descriptores morfológicos.

Contenido

Agradecimientos	IV
Presentación	V
Variabilidad Genética y Caracterización de Especies Vegetales <i>Rigoberto Hidalgo</i>	2
Caracterización Morfológica de Germoplasma. Estudios de casos	27
Caso 1. Análisis de la variabilidad genética en quinua <i>Wilfredo Rojas</i>	27
Caso 2. Análisis de la variabilidad genética en frijol <i>Gustavo Ligarreto</i>	40
Caso 3. Variabilidad genética de jícama <i>César Tapia</i>	50
Caso 4. Análisis de divergencias interespecíficas con pasifloras andinas <i>Sergio Segura</i>	56
Conceptos y Mediciones Útiles para la Caracterización de Germoplasma	72
Análisis estadístico de datos de caracterización morfológica <i>J. L. Chávez</i>	72
Caracterización de germoplasma <i>Gustavo Ligarreto</i>	77
Divergencias morfológicas interespecíficas del subgénero <i>Tacsonia (Passiflora)</i> <i>Sergio Segura</i>	79
Análisis multivariado en estudios de variabilidad genética <i>Wilfredo Rojas</i>	85
Bibliografía Consultada	87

Variabilidad Genética y Caracterización de Especies Vegetales

*Rigoberto Hidalgo**

Como sucede con todos los organismos vivos que se desarrollan en condiciones naturales, la población de individuos que conforman una especie vegetal están bajo una continua interacción dinámica de adaptación con los factores en los que crece esa población. Dichos factores son los bióticos (microorganismos, otras especies vegetales, animales inferiores y superiores) y los abióticos (clima y suelo), para ello, cada especie adapta la información contenida en el genoma de acuerdo con las necesidades de sobrevivir en su entorno. El resultado de esta interacción adaptativa se traduce en la acumulación de la información genética que a manera de variantes cada especie va guardando entre los miembros de su población, y que se va transmitiendo en las subsiguientes generaciones a través del tiempo. De esta manera, aunque la población de individuos en una especie comparte características comunes y se pueden cruzar entre ellos, también es cierto que en cada uno existen muchas variantes individuales. La suma de todos los individuos con sus respectivas variantes es lo que se conoce como variabilidad genética de una especie, la cual permite a dicha especie adaptarse a los cambios que se pueden presentar en su entorno.

Fuentes de variabilidad

Existen numerosos tratados en los que se discute cómo se ha producido y aún se produce la variabilidad de las especies vegetales. Sin embargo, para los propósitos prácticos de este boletín, las fuentes de variabilidad para las especies de plantas cultivadas se pueden resumir en las categorías siguientes.

Evolutiva

Se refiere a la variabilidad producida durante los procesos evolutivos de especiación por los que haya pasado una especie, principalmente durante las etapas de aislamiento reproductivo, así como a la dinámica que la especie ha tenido y sigue teniendo en condiciones naturales. En este aspecto Ford-Lloyd y Jackson (1986) consideran que los patrones de diversidad genética de las plantas cultivadas resultan de la interacción de los factores principales siguientes: mutación, migración, recombinación, selección (natural y artificial) y deriva genética. Los tres primeros estimulan la producción de nueva variabilidad, mientras que los dos restantes pueden reducirla. En esta interacción entra a jugar un papel relevante la biología reproductiva –autógama o alógama, con sus variantes– que desarrolle la especie esperando, por lo general, una mayor variabilidad en las alógamas que en las autógamas.

Geográfica

Esta fuente de variabilidad es importante para un buen número de especies cultivadas que tienen un amplio rango de distribución geográfica, porque además de su dispersión natural, han sufrido una extensa dispersión artificial por acción del hombre. En ambos

* Ing. Agr. Especialista en Recursos Fitogenéticos.

casos, al llegar a un nuevo nicho ecológico empiezan un nuevo proceso evolutivo en el cual crean variantes genéticas de adaptación como respuesta a variaciones en los componentes ambientales. Una vez más entran a jugar los factores principales mencionados en la variabilidad evolutiva. En términos generales, se espera que a mayor rango de dispersión geográfica de una especie vegetal, ocurra una mayor variabilidad.

Domesticación

Durante el proceso de domesticación de las especies cultivadas el hombre ha ejercido una fuerte presión de selección que ha permitido la preservación de muchas variantes las cuales, posiblemente, hubieran desaparecido en condiciones naturales. De la misma manera, el hombre también indujo la producción de nuevas variantes, tanto para facilitar el manejo agronómico como para incrementar la producción. Este fenómeno se puede encontrar en todas las especies cultivadas, especialmente en las altamente domesticadas como los cereales (trigo, maíz, arroz), papa, frijol y otras. En el proceso de domesticación se pueden identificar dos etapas distintas de presión de selección del hombre con el objeto de preservar o producir variabilidad: (1) La domesticación que abarca todo el proceso de selección empírica mediante el cual el hombre fue adaptando las especies para suplir sus necesidades básicas en alimentación, vestido, salud e industria. Esto fue posible mediante la selección y conservación de variantes útiles que aparecían en las poblaciones en un proceso que para la mayoría de las especies cultivadas tuvo una duración superior a 10,000 años. (2) El descubrimiento de la genética que proporcionó una vía alterna a la naturaleza permitiéndole ampliar su variabilidad en el corto tiempo. Por otra parte, desde el comienzo del siglo XX, es decir hace aproximadamente 100 años, genetistas y mejoradores han estado produciendo nuevas variantes genéticas mediante infinidad de cruzamientos en la búsqueda de solucionar problemas de producción y aquellos ocasionados por plagas y enfermedades en las especies cultivadas.

Expresión de la variabilidad

Toda la variabilidad producida en los procesos descritos anteriormente se almacena en el genoma, es decir, entre los miembros de la población que conforman la especie, y puede o no expresarse en características que permitan ser identificadas. Por tanto, desde el punto de vista de su expresión, la variabilidad contenida en el genoma de una especie puede ser agrupada en dos grandes clases: (1) la que se expresa en características visibles y que conforman el fenotipo, y (2) la que no se expresa en características visibles y que en general se refiere a los procesos o productos internos de la planta. Sin embargo, vale la pena resaltar que muchos procesos en esta última clase están siendo identificados mediante técnicas de biología molecular que aún no son rutinarias en los bancos de germoplasma, pero que se espera lo sean en el futuro cercano. Es necesario distinguir entre lo que puede o no ser expresado en forma visual, con el fin de precisar qué porción de la variabilidad total de la especie se está analizando en la caracterización.

En relación con el fenotipo, los caracteres que lo conforman corresponden en su gran mayoría a la descripción morfológica de la planta y su arquitectura. Estos caracteres se denominan descriptores morfológicos y se pueden agrupar en los tipos que aparecen a continuación.

Botánicos-taxonómicos

Corresponden a los caracteres morfológicos que describen e identifican la especie y son comunes a todos los individuos de esa especie. En su gran mayoría estos caracteres tienen una alta heredabilidad y presentan poca variabilidad, aunque en las especies cultivadas con frecuencia se pueden encontrar unos pocos que muestran diferentes grados de variabilidad, especialmente en aquellos de interés particular para el hombre como son el tipo y la forma de la hoja, la forma del fruto y la descripción de la flor.

Morfoagronómicos

Corresponden a los caracteres morfológicos que son relevantes en la utilización de las especies cultivadas. Pueden ser de tipo cualitativo o cuantitativo, e incluyen algunos de los caracteres botánicos-taxonómicos más otros que no necesariamente identifican la especie, pero que son importantes desde el punto de vista de necesidades agronómicas, de mejoramiento genético, y de mercadeo y consumo. A manera de ejemplos de estos caracteres se puede mencionar la forma de las hojas; pigmentaciones en raíz, tallo, hojas y flores; color, forma y brillo en semillas; tamaño, forma y color de frutos; arquitectura de planta expresada en hábito de crecimiento y tipos de ramificación. Algunos curadores de bancos de germoplasma incluyen descriptores relacionados con componentes de rendimiento con el objetivo de proveer a los fitomejoradores indicación del potencial de este carácter en el germoplasma conservado. En su gran mayoría, estos descriptores tienen aceptable heredabilidad local pero son afectados por cambios ambientales.

Evaluativos

Esta porción de la variabilidad sólo se expresa como respuesta a estímulos ambientales bióticos (plagas y enfermedades) o abióticos (estrés por temperatura, agua, nutrientes). En general, la respuesta se expresa en características de tipo cualitativo.

El estudio de la variabilidad no expresada en características visibles en la planta se concentra en la detección de marcadores moleculares, entre los que se incluyen proteínas, isoenzimas y fragmentos de ADN, tema que se discutirá más adelante. Esta variabilidad se puede detectar y cuantificar mediante técnicas de biología molecular que todavía están en proceso de refinamiento. El continuo desarrollo de estas técnicas está aportando innumerables herramientas para el entendimiento más profundo de la evolución y la variabilidad genética y su utilización y, finalmente, para el mapeo genético de especies.

En resumen, existe una alta variabilidad genética en las especies vegetales como resultado de su respuesta para adaptarse a los cambios y presiones de los medios biótico y abiótico que las rodea. La suma de todas esas respuestas, es decir, de todos los miembros de la población, conforma la variabilidad genética de la especie.

La información genética de esa variabilidad se conserva y transmite por generaciones a través de los miembros de la población de la especie. Aunque dicha información se mantiene mediante una dinámica continua entre los miembros de la especie, la expresión de esa variabilidad puede o no manifestarse en caracteres visibles. La variabilidad que se expresa en caracteres visibles se denomina fenotípica y dentro de ella se encuentran las características botánicas-taxonómicas, las morfoagronómicas y las evaluativas como respuesta a factores bióticos y abióticos. La variabilidad que no se expresa en características requiere para su identificación el uso de técnicas especiales de laboratorio que en la actualidad se refieren principalmente a marcadores moleculares. La variabilidad es más

evidente en las especies vegetales cultivadas, porque adicionalmente a las presiones del medio, han sufrido la selección ejercida por el hombre para adaptarlas a sus propósitos.

El interés por la variabilidad de las especies vegetales

Antecedentes históricos

El hombre a través de la historia ha dependido de las plantas para su supervivencia. Como se mencionó antes, las plantas en su estado natural tienen una dinámica evolutiva y están continuamente produciendo variabilidad. Tanto la variabilidad visible como la no-visible han sido usadas por el hombre para identificar, estudiar y utilizar las especies vegetales. A continuación se presenta una síntesis de los principales eventos históricos a través de los cuales el hombre se interesó progresivamente por la variabilidad existente en las especies vegetales.

Primeros recolectores. Las características morfológicas de las plantas han sido utilizadas por el hombre desde el momento en el cual comenzó a recolectar semillas y a seleccionar especies vegetales que le podían servir para satisfacer sus necesidades básicas. Así, mediante la identificación de características claves como colores, formas, olores y texturas le fue posible inferir sobre los usos potenciales de una especie en particular. El proceso de domesticación de la gran mayoría de las especies cultivadas que hoy se conocen tuvo una duración superior a 10,000 años y durante ese tiempo se acumuló una gran cantidad de variantes genéticas en cada especie, las cuales es posible diferenciar en forma visual por sus características fenotípicas.

Botánica clásica. A medida que se fueron seleccionando cada vez más especies útiles y muchas de sus variantes genéticas, se acumuló información valiosa que fue transmitida por generaciones, especialmente en aquellas culturas establecidas en los centros de origen y domesticación de cultivos. Sin embargo, aunque esa información era sumamente útil desde la perspectiva de utilización, su importancia se tornó más relevante cuando en el siglo XVIII aparecieron los botánicos clásicos encabezados por Carlos Lineo, quien inició el gigantesco trabajo de clasificar científicamente las especies vegetales, dando inicio al uso de la nomenclatura binomial para identificarlas. El éxito de esta tarea se fundamentó en el conocimiento detallado de las características morfológicas de todas las partes componentes de la planta, especialmente de la flor.

Además de la clasificación individual de las especies, otros investigadores como De Candolle, Bentham, Hooker, Engler y Prantl iniciaron una tarea paralela de clasificación filogenética y taxonómica de grupos superiores como órdenes, familias, tribus y géneros mediante la identificación de características comunes a esos grupos (Hutchinson, 1964). Más tarde, a finales del siglo XIX, se comenzó a elucidar la distribución geográfica de los grandes grupos botánicos. Nació, entonces, la geografía botánica encabezada por Alexander von Humboldt (Plucknett et al., 1992) en la cual los datos relacionados con los sitios de recolección empezaron a ser importantes, dando origen a lo que posteriormente se denominaría 'datos pasaporte'.

Genética-centros de origen. Desde mediados del siglo XIX han ocurrido varios acontecimientos que resaltan, aún más, la importancia de la variabilidad genética

producida durante siglos en los centros geográficos de diversidad de cultivos. En este período, además de identificar la variabilidad de las características de las plantas, también se estableció la posibilidad de manipular dichas características. Entre los eventos sobresalientes se pueden mencionar los siguientes:

Descubrimiento de la genética y su utilidad. Los postulados de Mendel (1865) fueron redescubiertos en el comienzo del siglo XX. Gracias a este acontecimiento se pudo entender cómo se transmiten las características de las plantas de una generación a otra abriendo, así, la genética un potencial inmenso como herramienta útil en la manipulación de dichas características, es decir, de la variabilidad.

Necesidad de producir más alimentos. Al comienzo del siglo XX ya era apremiante la necesidad de producir más alimentos para suplir la demanda de la creciente población mundial. Con la adopción de la genética como herramienta de manipulación aparecieron los mejoradores de plantas, quienes, sin embargo, para poder cumplir con sus objetivos necesitaban disponer de una mayor variabilidad de especies cultivadas.

Teoría sobre los centros de origen de los cultivos. En la década de 1920-30 el botánico ruso N. I. Vavilov concibió métodos detallados de clasificación basados en las características de las plantas, entre ellos: especies, subespecies, variedad, subvariedad y forma, y los correlacionó con la distribución geográfica para determinar dónde se encontraba la mayor variabilidad. Con esa información propuso su teoría de los centros de origen de las especies cultivadas, la cual ayudó a los fitomejoradores a encontrar la diversidad que estaban buscando (Hawkes, 1970).

Establecimiento de bancos de germoplasma. Con el conocimiento más detallado de los centros de origen, varios países empezaron a recolectar la variabilidad de los cultivos existente en ellos y surgió, entonces, la necesidad de establecer los bancos de germoplasma. La FAO desde 1960 inició la tarea de concientizar a las instituciones sobre la necesidad de preservar los recursos genéticos para, por un lado, tener herramientas con el fin de aliviar la cada vez más grave escasez de alimentos y, por otro, salvar la variabilidad de aquellas especies cultivadas que se encuentran en peligro de erosión en los centros de origen y de diversidad. Así nacieron los bancos de germoplasma en los centros internacionales de investigación agrícola.

ADN, marcadores moleculares y biología molecular. A mediados del siglo XX Watson y Crick (1953) descubrieron la estructura del ADN lo que permite el conocimiento básico del código genético y los mecanismos mediante los cuales la totalidad de características visibles y no-visibles de un organismo se establecen y son transmitidas de generación en generación. Alrededor del estudio del ADN se desencadenó, entonces, una serie de nuevas áreas de investigación que en las últimas dos décadas se han fusionado bajo el común denominador de biología molecular. Son innumerables las consecuencias de este descubrimiento para la evolución, la biología y la genética. Es importante, también, resaltar que así como los factores bióticos y abióticos estimulan nueva variabilidad visible y aun pueden modificar la ya existente, igualmente pueden producir cambios en moléculas básicas constitutivas del núcleo y del citoplasma de una célula. Como resultado surgió el concepto de marcadores moleculares o tipos

adicionales de características relacionadas con los componentes moleculares y el ADN constitutivos de las células y cuyo uso cada día es mayor para describir la variabilidad del germoplasma de una especie. La variabilidad o polimorfismo de estos marcadores puede ser usado en los diferentes estudios sobre la diversidad genética de las especies vegetales y su herencia puede ser monitoreada. Los grandes grupos de marcadores moleculares reconocidos son: (1) las proteínas, que pueden ser de reserva o almacenamiento, y las isoenzimas o aloenzimas que son formas diferentes de una enzima que comparten una misma actividad catalítica y pueden ser denominadas como isoenzimas; y (2) el ADN cuyas diferencias en las secuencias, principalmente de fragmentos, se presentan por poliformismos o variabilidad.

En resumen, la secuencia histórica antes descrita muestra cómo el hombre ha identificado, conservado y utilizado la variabilidad desarrollada por las especies vegetales como respuesta de adaptación y supervivencia a los cambios que presentan los medios bióticos y abióticos que las rodean. Esa variabilidad es notoria en las especies cultivadas y fue la base para la supervivencia de las primeras civilizaciones, el desarrollo de la botánica y la taxonomía, el descubrimiento de la genética, del desarrollo de la teoría de los centros de origen y el establecimiento de los bancos de germoplasma.

Caracterización de la variabilidad

En la caracterización de una especie se estima la variabilidad existente en el genoma de la población de individuos que la conforman. Así, el genoma de las especies de animales o plantas contiene toda la información codificada en forma de genes que se necesitan tanto para establecer su identidad morfológica como para desarrollar todos los procesos y funciones vitales para su supervivencia. Se estima que las plantas superiores poseen un poco más de 400,000 genes con funciones particulares dentro de la especie y un buen número de ellos ha creado variantes por efectos evolutivos y del medio ambiente. Esas variantes se van acumulando entre los diferentes miembros componentes de la especie y la suma de todos los efectos de los genes y sus variantes es lo que se denomina variabilidad genética de una especie.

Anteriormente se mencionó que los todos los genes cumplen determinadas funciones y sus efectos pueden o no expresarse en características identificables de forma visual. Esto quiere decir que hay una variabilidad que se puede detectar a simple vista y otra que, aunque no es visible fácilmente, también existe en la especie pero que requiere de técnicas especiales para ser detectada. Por ello, es primordial identificar cuál es el nivel de variabilidad que se intenta medir o describir con el fin de elegir las herramientas o métodos estadísticos adecuados para analizar los datos resultantes de un estudio de caracterización.

El primer nivel se refiere a la caracterización de la variabilidad detectable visualmente, la cual se puede dividir en los tipos siguientes: (1) Las características responsables de la morfología y la arquitectura de la planta utilizadas en un principio para la clasificación botánica y taxonómica, aunque en muchas de ellas se pueden encontrar variantes. (2) Una serie de características relacionadas especialmente con aspectos de manejo agronómico y de producción de la especie que son de interés para mejoradores y agrónomos. En la mayoría de los bancos de germoplasma de programas existentes actualmente se hace una caracterización morfoagronómica en la que se fusionan estos dos primeros tipos. (3) Un grupo de características detectables visualmente que

sólo se expresan como reacción a estímulos del medio ambiente. Estos pueden ser biótico como plagas y enfermedades; o abióticos como sequías, deficiencias de minerales y cambios en temperatura, entre otros. Este tipo de caracterización se denomina evaluación y para su correcta cuantificación, generalmente, se requieren diseños experimentales separados de la caracterización morfoagronómica.

El segundo nivel se refiere a la caracterización de la variabilidad que no es detectable por simple observación visual. Esta caracterización se denomina molecular porque se refiere a la identificación de productos y/o funciones internas de la célula. Todas las técnicas de laboratorio para detectar esta variabilidad se agrupan dentro del concepto de marcadores moleculares explicado anteriormente. Si bien ya existen algunos métodos de análisis de datos en proceso de desarrollo para estos tipos de caracterización las técnicas de laboratorio son relativamente recientes y están en continuo proceso de mejoramiento y actualización.

En este boletín se incluyen principalmente los métodos de análisis más comunes para el primer nivel que se refiere a la caracterización morfoagronómica, aunque dependiendo de las circunstancias, los datos de tipo evaluación también se pueden incluir dentro de este análisis. Para las especies vegetales cultivadas son de gran utilidad estos dos tipos de caracterización, ya que proporcionan una idea clara sobre la variabilidad de interés para los investigadores agrícolas y los curadores de las colecciones de germoplasma. En relación con el segundo nivel es necesario aclarar que, aunque algunos de los métodos de análisis de datos morfoagronómicos son utilizados para la caracterización molecular, este análisis no se considera en este boletín. Sin embargo, es oportuno enfatizar nuevamente que la actual tendencia consiste en integrar globalmente los resultados de los diferentes tipos de caracterización (morfoagronómica + evaluación + molecular) en forma complementaria con datos de información geográfica de origen lo que permite tener una idea más completa en el intento de estimar la variabilidad total del genoma de una especie vegetal. Este último objetivo empieza a ser prioritario en los bancos de germoplasma nacionales e internacionales.

Por tanto, el enfoque actual del estudio de la variabilidad genética de las colecciones de germoplasma de una especie exige la documentación lo más completa posible de las accesiones componentes de dichas colecciones. Esta documentación incluye información sobre el origen geográfico (localización geográfica, altitud, clima y suelo), la caracterización morfológica, la evaluación de la respuesta a factores bióticos y abióticos y, por último, una caracterización de marcadores moleculares (proteínas, enzimas, ADN) para cada accesión. Estos grupos de información, aparentemente independientes, se encadenan para finalmente establecer criterios racionales que permiten explicar la variabilidad de la especie en estudio. No obstante, es necesario enfatizar que independientemente del nivel de caracterización, los resultados del análisis de los datos son sólo una estimación de la variabilidad total de la especie.

Objetivos de la caracterización

En el proceso de caracterización de una colección, independientemente de su tamaño, se pueden establecer los objetivos principales siguientes: (1) Medir la variabilidad genética del grupo en estudio; para lo cual se pueden incluir uno, varios o todos los niveles posibles de variabilidad, es decir, fenotípica, evaluativa y molecular, utilizando en todas ellas descriptores previamente definidos. (2) Establecer la representatividad de la colección y

su relación con la variabilidad de la especie en una región, o con la variabilidad total de la especie. (3) Investigar la estructura genética, o sea, la forma como se compone la colección estudiada en relación con las variantes, o sus combinaciones que forman grupos o poblaciones identificables. Lo anterior está influenciado por factores demográficos in situ, tales como tamaño de población, biología reproductiva y migración. (4) Identificar los porcentajes de duplicidad de accesiones que puedan existir en una misma colección o en comparación con otras colecciones de la especie. (5) Identificar genes especiales o alelos particulares que pueden ser de carácter individual o en combinaciones únicas y que se pueden expresar en caracteres visibles (morfológicos o de evaluación) en diferentes estados o combinaciones de estados. A estos genes generalmente se les denomina 'stocks genéticos' y son utilizados para investigaciones de aplicación práctica inmediata, como es el caso de resistencia a factores bióticos.

En este boletín se presentan las metodologías y los análisis más usados y disponibles para medir la variabilidad genética que se expresa en características visibles, especialmente la caracterización morfológica y la evaluación. En un futuro se espera tratar los temas relacionados con el análisis integrado utilizando información geográfica y marcadores moleculares.

Recomendaciones prácticas para la caracterización

La caracterización de la variabilidad genética tiene varias limitaciones que son comunes a la gran mayoría de los bancos de germoplasma en instituciones nacionales e internacionales, y que se deben tener en cuenta en el momento de planificar los procedimientos. La primera es la escasa cantidad y la baja calidad de las semillas disponibles, lo que no permite flexibilidad en los trabajos de multiplicación y caracterización. La segunda es la pobre documentación de las colecciones, debido, principalmente, a que un alto porcentaje del germoplasma existente en los bancos ha sido producto de recolecciones oportunistas que no fueron hechas con un criterio de recursos genéticos y, más bien, en la búsqueda de características agronómicas. Finalmente, existe una baja disponibilidad de recursos para el mantenimiento sostenible de los bancos de germoplasma, lo que se refleja en un bajo número de accesiones y en una reducida cantidad de semillas por accesión.

En consecuencia cuando se caracteriza una colección de germoplasma se sugiere tener en cuenta las recomendaciones siguientes:

- Es necesario tener un conocimiento completo de la biología de la especie, especialmente en el aspecto reproductivo –sexual, asexual, autógama, alógama– así como de su centro de origen y domesticación.
 - La documentación adecuada proporciona elementos útiles para establecer una visión preliminar de la colección en referencia. Con esa visión es posible inferir de la variabilidad que se puede encontrar en los materiales aún antes de iniciar la caracterización, e igualmente, ayuda en la definición clara de los objetivos de la caracterización y permite ahorrar pasos innecesarios.
 - Se deben establecer claramente los objetivos teniendo en cuenta si lo que se busca es la variabilidad del grupo, la representatividad de la colección, investigar la estructura, identificar duplicados o detectar genes especiales.
-

- Independiente de los objetivos establecidos, se recomienda realizar una siembra experimental previa que permita conocer en términos generales la variabilidad global de la colección, la facilidad de registro de los descriptores y su utilidad para la caracterización y multiplicación de semillas.
- Antes de intentar la caracterización definitiva se recomienda homogenizar las accesiones de acuerdo con sus morfotipos. Esto es especialmente importante con formas silvestres y variedades tradicionales nativas, las cuales frecuentemente en su estado original son mezclas de morfotipos, por ej., tipos de semilla, hábitos de crecimiento, colores de flor y tipos de frutos. Aunque en el banco de germoplasma se conserve una muestra completa del original, la caracterización de una accesión que tiene mezcla de morfotipos dificulta enormemente el análisis de datos. Si no se puede hacer la homogenización al momento de preparar las semillas se debe intentar, hasta donde sea posible, hacer una siembra experimental previa para lograr dicho propósito.
- Para obtener mejor y mayor información en el análisis estadístico y confiabilidad en las diferencias entre los materiales y las variables se recomienda establecer 3 a 5 plantas de cada accesión y un mínimo de dos replicaciones.
- Cuando la disponibilidad de semillas o de material vegetativo es baja y, por tanto, no es posible establecer parcelas replicadas de cada accesión, se recomienda seleccionar un lote lo más homogéneo posible para evitar los efectos de la variabilidad en las condiciones del suelo. En estos casos la obtención correcta de los datos facilita el análisis comparativo entre accesiones y aun entre variables.
- Si el objetivo principal es medir la variabilidad del grupo, se recomienda seleccionar descriptores que sean lo más discriminatorios posible. Esto permite ahorrar tiempo al evitar la toma de datos repetitivos y simplifica el análisis. En consecuencia, se sugiere consultar los listados de descriptores publicados por el IPGRI para la especie en estudio.
- En el momento de diseñar el trabajo de caracterización se recomienda consultar con un estadístico o un profesional afín sobre el diseño en el campo, la forma adecuada de registrar y analizar los datos, y la interpretación de los resultados.
- El uso de los programas automatizados actualmente disponibles ayuda en el entendimiento de los procedimientos relacionados con los métodos estadísticos avanzados para el análisis de datos de caracterización, especialmente los multivariados. La clave es saber interpretar los resultados en el punto donde el conocimiento biológico de la especie es importante para explicar los resultados del análisis de los datos.

En este boletín se hace referencia a las técnicas de análisis de datos más frecuentemente utilizados para la caracterización morfoagronómica de germoplasma de especies vegetales cultivadas. Por tanto, se recomienda consultar los estudios de casos que se incluyen en la siguiente sección y que han sido seleccionados tratando de incluir una gama amplia de especies cultivadas como fríjol, quinua, pasifloras y jícama. En cada caso se presentan diferentes enfoques que resultan útiles para entender los distintos tipos de análisis de datos.

Descriptores

Un descriptor es una característica o atributo cuya expresión es fácil de medir, registrar o evaluar y que hace referencia a la forma, estructura o comportamiento de una accesión. Los descriptores son aplicados en la caracterización y evaluación de las accesiones debido a que ayudan a su diferenciación y a expresar el atributo de manera precisa y uniforme, lo que simplifica la clasificación, el almacenamiento, la recuperación y el uso de los datos. Estos descriptores han sido definidos para un gran número de especies cultivadas. El IPGRI ha compilado y publicado en forma de manual listados de descriptores para más de 100 especies cultivadas. A continuación se incluyen los diferentes tipos de descriptores cuya definición completa aparece en las listas del IPGRI.

De pasaporte. Proporcionan la información básica que se utiliza para el manejo general de la accesión, incluyendo el registro en el banco de germoplasma y cualquier otra información de identificación, y describen los parámetros que se deben observar cuando se hace la recolección original.

De manejo. Proporcionan las bases para el manejo de las accesiones en el banco de germoplasma y ayudan durante su multiplicación y regeneración; por ej., fechas de multiplicación, cantidades de semillas disponibles, porcentajes de viabilidad.

Del sitio y el medio ambiente. Describen los parámetros específicos del sitio y del ambiente y ayudan en la interpretación de resultados cuando se realizan pruebas de caracterización y evaluación. Se incluyen, también, en esta categoría los descriptores del sitio de recolección del germoplasma; por ej., coordenadas geográficas, características de clima y suelos.

De caracterización. Permiten la discriminación relativamente fácil entre fenotipos. Generalmente son caracteres altamente heredables que pueden ser fácilmente detectados a simple vista y se expresan igualmente en todos los ambientes. Además, pueden incluir un número limitado de caracteres adicionales considerados como deseables por consenso de los usuarios de un cultivo en particular; por ej., colores y formas de tallos, hojas, flores, semillas y frutos. Adicionalmente, en los últimos años se están incluyendo descriptores relacionados con los marcadores moleculares, gracias a los avances logrados en la biología molecular, especialmente en las técnicas de electroforesis.

De evaluación. La expresión de la mayoría de los descriptores de esta categoría depende del medio ambiente y, en consecuencia, se requieren métodos experimentales especiales para su evaluación. La evaluación puede también involucrar métodos complejos de caracterización molecular o bioquímica. En este tipo de descriptores se incluyen caracteres como rendimiento, productividad agronómica, susceptibilidad a estrés y caracteres bioquímicos y citológicos, los cuales generalmente son de mayor interés en el mejoramiento de cultivos.

Para los propósitos de este boletín se dará énfasis a los descriptores de caracterización y de evaluación, aunque es importante resaltar nuevamente que todos ellos tienen un papel relevante en el análisis integral de una colección de germoplasma.

Estados del descriptor y tipos de datos

Se espera que las características visibles de una especie sean más o menos homogéneas, sin embargo, todas no se expresan con la misma intensidad y algunos miembros de la población pueden presentar diferentes grados de expresión que se traducen en diferentes tipos de datos o categorías de variables. Por tanto, los descriptores se pueden diferenciar de acuerdo con el estado que presentan, lo cual es conocido como 'estados del descriptor' y se registran mediante escalas de valor.

Existen distintas categorías de datos, según la expresión del descriptor que puede ser en forma cualitativa o cuantitativa (Figura 1). Si se expresa en forma cualitativa, se pueden generar datos binarios (también llamados de doble estado), datos con secuencia (ordinales) y datos sin secuencia (nominales). Si se expresa en forma cuantitativa, los datos generados pueden ser discontinuos (llamados también discretos) y continuos. Las siguientes sugerencias ayudan en el registro práctico de los datos:

- Para los datos cualitativos de tipo binario, cada descriptor presenta dos estados (presente = 1, ausente = 0). Por ejemplo, presencia de flores blancas (1), ausencia de flores blancas (0).
- Para los datos cuantitativos de tipo ordinal o con secuencia, el descriptor se registra utilizando una serie de estados predefinidos; por ejemplo, para altura de la planta: 1 = corta (<0.5 m), 3 = intermedia (>0.5 <1.5 m), 5 = alta (>1.5 m).
- Para los datos cualitativos de tipo nominal o sin secuencia el descriptor se registra usando una serie de estados previamente definidos; por ejemplo, 1 = blanco, 2 = crema, 3 = amarillo.
- Para los datos cuantitativos de tipo continuo el descriptor se registra en unidades internacionales (SI) estándar, por ejemplo, altura de la planta = 0.9 m; peso de 100 semillas = 250 g.

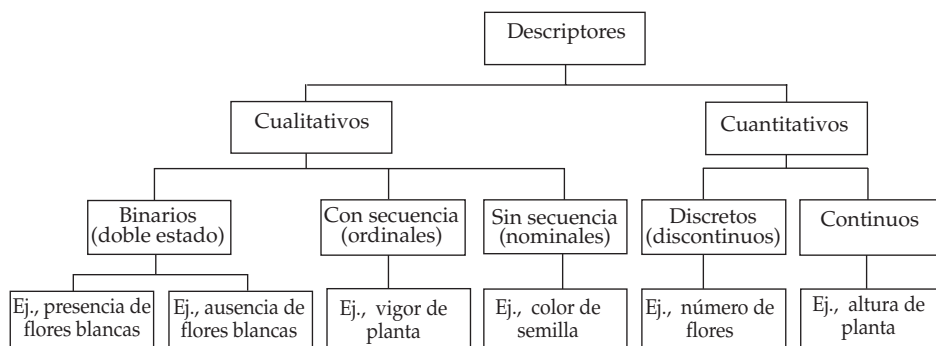


Figura 1. Tipos de datos y descriptores

Organización de los datos, matriz básica de datos

La matriz básica de datos (MBD) se construye a partir de la información que se obtiene en la caracterización y evaluación de especies. Consiste en un arreglo en forma de cuadrícula con tantas filas como accesiones existentes (**n**) y una columna para cada variable (**p**) (Figura 2).

		Variables			
		X1	X2	X3	X4... Xp
Accesiones	A1				
	A2				
	A3				
	A4				
	An				

n accesiones medidas sobre p variables

Figura 2. Esquema de una matriz básica de datos (MBD).

La construcción de la MBD es fundamental porque constituye el punto de partida o materia prima para la aplicación de las herramientas estadísticas. Desde el punto de vista geométrico y espacial la matriz $n \times p$ puede ser conceptualizada de las formas siguientes (Pla, 1986): (1) Como un conjunto de n accesiones en un espacio definido por las p variables, donde las observaciones serán puntos que representan a accesiones en el espacio. Las accesiones consideradas se comparan en función de sus variables. (2) Como un conjunto de p variables, donde las observaciones corresponderán a puntos que representan las variables. Las variables consideradas se comparan en función de las accesiones. En el cuadro 1 se presenta un ejemplo de la caracterización de cuatro accesiones y tres variables del cultivo de frijol (*Phaseolus vulgaris*), utilizando una MBD.

Cuadro 1. Ejemplo de una matriz básica de datos (MBD) para características del cultivo de frijol.

Accesión (no.)	Días a floración (no.)	Peso de 100 semillas (g)	Altura de planta (m)
1	30	25	1.06
2	37	32	0.34
3	45	60	1.60
4	25	24	0.86

Tamaño de la muestra

Luego de definir los datos para caracterizar los genotipos mediante descriptores elaborados por el IPGRI o desarrollados por los investigadores, el siguiente paso consiste en estimar el tamaño óptimo de la muestra o número de observaciones para que el parámetro represente la muestra del cultivar en estudio. Para obtener estos datos, el investigador debe hacer estudios previos en los que se presente la mayor variabilidad posible de la especie; de esta manera se garantiza que la muestra no cambie sensiblemente de una población a otra (Enríquez, 1991). Si el estudio se basa en una población homogénea se corre el riesgo de subestimar el tamaño de la muestra, razón por la cual es necesario garantizar la mayor variabilidad posible (Enríquez, 1991).

Para calcular el tamaño de la muestra se sugiere la metodología siguiente que ha sido utilizada con éxito por la sección de Mejoramiento de la Universidad Nacional de Colombia, seccional Palmira. Dicha metodología se basa en la ecuación:

$$n = \frac{4CV^2}{E_{\%}^2} \quad (1)$$

donde,

CV = Porcentaje de variación asociado con el descriptor que se considere más variable dentro de la colección. Este valor se puede obtener de investigaciones previas o en la literatura.

$E_{\%}^2$ = Error permisible expresado como porcentaje de la media verdadera. Se refiere a la diferencia que se espera entre la media muestral \bar{x} y la media verdadera μ del descriptor, expresada como porcentaje de la media verdadera (μ) con un nivel de confianza de 95%. En el Cuadro 2 se incluyen algunas opciones de tamaños de muestra para diferentes combinaciones CV y $E_{\%}^2$.

En el ejemplo del Cuadro 2 se requieren como mínimo 16 plantas para alcanzar un error permisible de 20% en un descriptor con un CV aproximado de 40%. El tamaño de la muestra en muchos casos es limitado por la cantidad de semillas disponible, lo que no permite hacer estimaciones más precisas del promedio real de un descriptor en las colecciones. Si el tamaño de la muestra es mayor que 4 plantas por accesión, éstas se deben disponer en varias parcelas o repeticiones con 2 ó 3 plantas por parcela.

Cuadro 2. **Tamaño de muestra para diferentes combinaciones de CV y $E_{\%}^2$, utilizando la ecuación 1.**

$E_{\%}^2$	Coeficientes de variación (CV, %)						
	10	15	20	25	30	35	40
10	4	9	16	25	36	49	64
15	2	4	7	11	16	22	28
20	1	2	4	6	9	12	16

Cuando es necesario hacer varias determinaciones de un mismo descriptor por cada planta, se sugiere realizar un muestreo preliminar para comparar la magnitud de la variación entre plantas con la variación entre determinaciones por planta, lo que permite establecer un método apropiado de muestreo.

Existen otras alternativas para calcular el número de observaciones, entre las cuales sobresale la propuesta por Steel y Torrie (1980). Para estimar el tamaño de muestra y realizar la programación que permita la aplicación de la fórmula anterior se puede utilizar el programa estadístico SAS u otro similar.

Análisis de datos

Métodos para el análisis de datos de caracterización. Los datos se pueden analizar mediante el empleo de métodos simples o complejos, que van desde el uso de gráficos y estadísticos de tendencia central y dispersión hasta los multivariados. El análisis tiene el propósito de reducir el volumen de información característico en trabajos de esta naturaleza.

Mediante la aplicación de estos métodos sobre la MBD es posible obtener conclusiones acerca de la variabilidad y la utilidad del germoplasma, por tanto, los datos deben

representar fielmente las características y el comportamiento de las accesiones.

Estadísticos simples. Permiten estimar y describir el comportamiento de las diferentes accesiones en relación con cada carácter. Los más comunes son el promedio, la media aritmética, el rango de variación, la desviación estándar (DE) y el coeficiente de variación (CV), que se utilizan en el análisis de datos cuantitativos. Estos se deben realizar antes de cualquier análisis multivariado, ya que proporcionan una idea general de la variabilidad del germoplasma y permiten inmediatamente detectar datos no esperados y errores de medición en el ingreso de datos, entre otros.

La media aritmética. Es una medida de tendencia central que ayuda a caracterizar el germoplasma y permite relacionar un atributo de una accesión con un valor central de dicho atributo.

El rango de variación o amplitud total. Se define como la diferencia entre el valor mínimo y el máximo de cualquier variable sobre el conjunto de accesiones estudiadas.

La desviación estándar. Cuantifica la magnitud de la variación respecto a la media aritmética y se expresa en las mismas unidades que las observaciones originales. Proporciona una idea del estado (próximas o dispersas) de la mayoría de las accesiones de la colección en relación con una característica considerada.

El coeficiente de variación. Es una medida relativa de variación que define más intrínsecamente la magnitud de la variabilidad de los caracteres estudiados debido a que es independiente de las unidades de medida. Facilita la comparación de la variabilidad de una misma característica en dos grupos de accesiones o de caracteres medidos sobre la misma colección.

En el Cuadro 3 se presenta un ejemplo del uso de los estadísticos simples aplicados a características cuantitativas de una colección de lulo (*Solanum quitoense*). Se observa que con dichos estadísticos se obtiene información útil que permite inferir varios resultados e interrogantes claves; por ej., los caracteres diámetro de tallo, número de frutos por planta y cosechados y el peso de estos tienen un $CV > 50\%$, lo cual sugiere que tienen la más alta variabilidad en la especie. Así mismo, aparecen seis variables con $CV < 20\%$, lo que indica que la especie puede tener poca variabilidad en estos caracteres. No obstante, el grado de variabilidad de un carácter no indica necesariamente la magnitud de su utilidad desde el punto de vista del cultivo, ya que esto depende de los usos de la especie.

Después de realizar el análisis exploratorio con el uso de estadísticos simples, se sugiere que antes de cualquier análisis posterior más complejo se evalúen aquellas variables que identifiquen grupos o subgrupos naturales de la especie, especialmente cualitativas como la forma y el color de las hojas, las flores y las semillas; y el hábito de crecimiento de las plantas. Este tipo de variables posiblemente no es objeto de análisis estadístico, pero pueden ayudar a entender o complementar los resultados finales del estudio de las demás variables. Para estos casos se sugiere elaborar las tablas de frecuencias con el fin de establecer las proporciones de los diferentes grupos dentro de una colección de germoplasma.

Cuadro 3. Descriptores^a de características cuantitativas y estadísticos simples en poblaciones de lulo (*Solanum quitoense*), estudiadas en el Valle del Cauca, Colombia. 1985-86.

Variables (descriptores)	Estadísticos				
	n	\bar{X}	s^2	r	CV (%)
Altura (cm)	88	103.7	453.9	39.7	20.5
Diámetro tallo (cm)	88	3.5	0.5	1.4	20.2
No. de chupones	88	2.9	2.5	3.0	54.5
Long. espinas en tallo (mm)	41	3.1	0.3	1.2	17.7
Long. peciolo (cm)	88	12.1	6.3	4.6	20.7
Long. máxima hoja (cm)	88	31.0	50.0	14.3	24.8
Ancho máximo hoja (cm)	88	27.4	58.3	13.9	27.9
Area foliar (cm ²)	71	640.4	67494.1	469.7	40.6
Long. espinas en hoja (mm)	41	2.9	0.6	1.7	26.7
No. venas/hoja	88	14.0	1.7	2.7	9.3
No. lóbulos repandos/hoja	88	12.3	3.5	3.6	15.2
No. interlóbulos/hoja	88	8.7	5.0	4.6	27.4
No. inflorescencias/planta	88	7.4	1.0	2.1	13.5
No. flores/inflorescencia	88	43.7	219.2	27.8	33.9
Long. eje principal de inflorescencia (cm)	88	2.4	0.02	0.4	5.9
Long. pedúnculo de fruto (cm)	88	1.4	0.02	0.3	10.1
No. de frutos/planta.	88	24.8	753.1	22.4	64.1
No. de frutos cosechados.	71	12.3	71.8	15.6	68.9
Peso frutos cosechados (g)	71	652.5	211 534.5	846.3	70.5

a. n = número de muestras. \bar{X} = media aritmética. S^2 = desviación estándar. r = rango de variación. CV = coeficiente de variación.

Medidas de similitud. Permiten conocer el grado de asociación entre las **n** accesiones o entre las **p** variables. Pueden ser de varias formas, aunque el índice de similitud y los coeficientes de correlación y distancia dominan las aplicaciones. Cada uno de ellos representa una perspectiva particular de similitud, dependiendo tanto de las accesiones como del tipo de datos.

El índice de similitud. Se recomienda para comparar accesiones cuyas características son evaluadas en una escala nominal o de datos de doble estado (presencia o ausencia) de las características medidas. Por lo general, los programas de informática dan un apoyo limitado a las medidas de asociación, lo que implica construir una tabla de doble entrada a partir de la cual se procede a calcular los coeficientes y con ellos a construir la matriz simétrica de similitud.

A continuación se introduce dicha matriz en los programas estadísticos (Crisci y López, 1983; Hair et al., 1992) para obtener resultados más precisos y útiles, tal como se muestra en el ejemplo de la Figura 3

Los valores que se obtienen de los coeficientes de similitud varían entre uno (1) y cero (0), siendo el valor 1 el de máxima similitud y el valor 0 el de mínima. Entre los índices más usados se tienen 'Simple Matching Coefficient' (SMC), Jaccard (CAJ), Rogers y Tanimoto (RT), y Dice o Sørensen (SD) (Crisci y López, 1983; Crivisqui, 1998).

El coeficiente de correlación. Cuantifica en términos relativos el grado de asociación íntima o variación conjunta entre dos descriptores cuantitativos, por ej., entre altura de planta y días a floración. Su valor oscila entre -1 y +1 (Cuadro 4). El signo del coeficiente

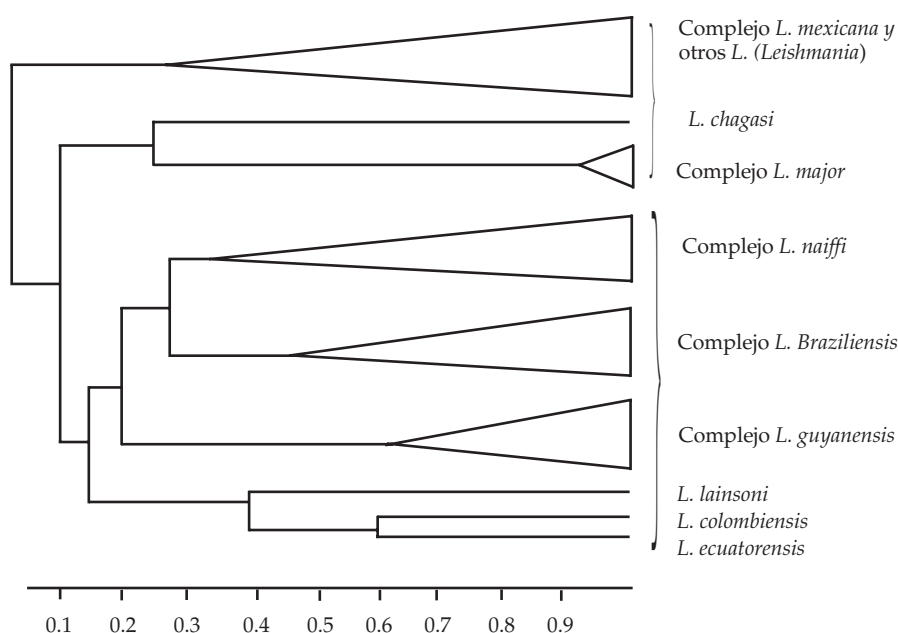


Figura 3. Dendrograma que muestra el nivel de similitud y la diversidad en cada complejo fenotípico de especies de *Leishmania*.

FUENTE: Cupolillo et al., 1998.

indica el tipo de asociación –negativo (-) si la relación es inversa y positivo (+) si es directa– La magnitud está asociada con el grado de intimidad entre las variables, si el valor es próximo a 1 están estrechamente correlacionadas; por el contrario, un valor próximo a 0 debe ser interpretado con reserva ya que puede indicar independencia entre las variables o una relación no lineal. El coeficiente más empleado es el de Pearson que se recomienda para datos de tipo multiestados cuantitativos, aunque también es útil para datos mixtos.

El coeficiente de distancia. Representa la similitud como la proximidad de las variables o accesiones con respecto a las demás. Son en realidad medidas de diferencias donde los valores elevados indican una menor similitud. Según Crisci y López (1983) su uso se recomienda para analizar datos cualitativos, cuantitativos y mixtos (ambos tipos). Los resultados se obtienen en una matriz simétrica cuyos valores varían de 0 a α (infinito), donde 0 es el indicativo de máxima similitud. Entre los coeficientes más usados se tiene 'Mean Character Difference' (MCD) y las distancias de Manhattan (MD), taxonómica (TD), euclidiana (DE) y euclidiana al cuadrado (d^2) (Crisci y López, 1983; Hair et al., 1992). En la Figura 4 se muestra un ejemplo de distancias para el caso de variables (descriptores) de quinua donde se observan variables fenológicas de arquitectura de la planta, grano, y emergencia de plántulas.

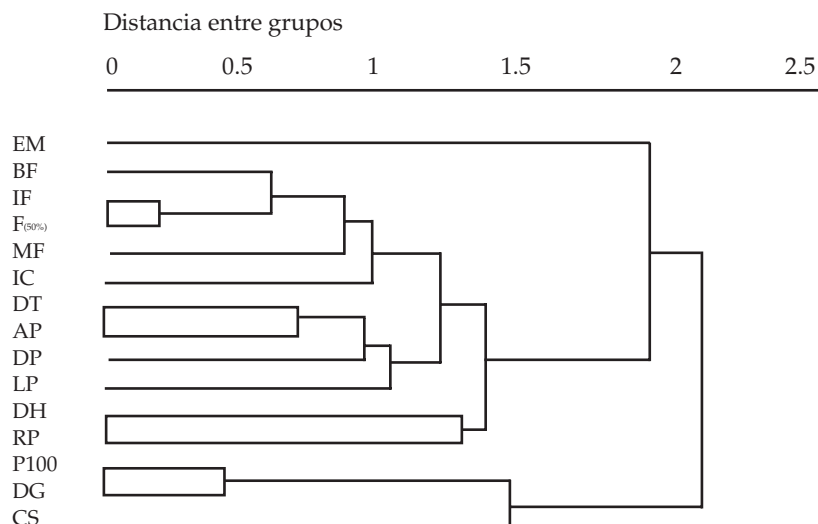
Como síntesis de los conceptos sobre los índices o coeficientes de similitud o de distancias, Bramardi (2002) explica que la semejanza entre pares de unidades ha recibido diferentes denominaciones tales como similitud, proximidad, disimilitud y distancia o

Cuadro 4. Ejemplo de una matriz de correlación simple entre 15 variables cuantitativas utilizadas para caracterizar germoplasma de quinua (n = 1512).

Características	EM	BF	IF	F _(50%)	MF	IC	RP	DH	DT	LP	DP	AP	DG	P100	CS
Emergencia	1.00														
Botón floral	0.14	1.00													
Inicio floración	0.21	0.69	1.00												
50% floración	0.16	0.73	0.94	1.00											
Madurez Fisiológica	0.25	0.37	0.63	0.61	1.00										
Indice cosecha	-0.09	-0.42	-0.59	-0.57	-0.55	1.00									
Rama principal	-0.01	0.37	0.39	0.44	0.16	-0.33	1.00								
Dientes hojas	0.05	0.38	0.43	0.45	0.23	-0.31	0.25	1.00							
Diámetro tallo	0.22	0.11	0.28	0.28	0.41	-0.20	0.31	0.12	1.00						
Longitud panoja	0.19	-0.30	-0.09	-0.15	0.34	0.02	-0.23	-0.24	0.40	1.00					
Diámetro panoja	0.12	-0.06	0.04	0.06	0.20	-0.02	0.20	0.01	0.60	0.31	1.00				
Altura planta	0.08	0.08	0.30	0.30	0.56	-0.32	0.27	0.10	0.69	0.58	0.38	1.00			
Diámetro grano	-0.21	0.03	-0.09	-0.07	-0.38	0.10	0.23	0.15	-0.17	-0.40	-0.05	-0.24	1.00		
Peso 100 granos	-0.22	-0.01	-0.16	-0.14	-0.42	0.17	0.14	0.15	-0.21	-0.36	-0.08	-0.25	0.89	1.00	
Cont. saponinas	-0.11	-0.13	0.12	0.16	-0.01	-0.11	0.21	0.23	0.01	-0.28	0.04	-0.08	0.40	0.34	1.00
EM		BF	IF	F _(50%)	MF	IC	RP	DH	DT	LP	DP	AP	DG	P100	CS

Valores en negrilla no son significativos (P ≥ 0.001).

FUENTE: Rojas, 1998.



EM = Emergencia, BF = Botón floral, IF = Inicio floración, F50% = 50% floración, MF = Madurez fisiológica, IC = Índice cosecha, DT = Diámetro tallo, AP = Altura planta, DP = Diámetro panoja, LP = Longitud panoja, DH = Dientes hojas, RP = Rama principal, P100 = Peso 100 granos, DG = Diámetro grano, CS = Contenido de saponinas.

Figura 4. Dendrograma de distancias entre 15 variables cuantitativas del germoplasma de quinua (usando la distancia euclidiana al cuadrado).

asociación, por tanto, se sugiere que todas estas acepciones sean consideradas como un grupo de índices de distancia. Así, por ejemplo, cuando se hace referencia a similitud se toma un rango entre 0 y 1, en el que 1 es similitud total y 0 es ausencia total de similitud. Por el contrario, cuando se refiere a distancia, aunque se toma el mismo rango, 0 significa similitud total mientras que 1 es ausencia total de similitud. La elección de cuál índice tomar depende de los datos y de los tipos de análisis que se planeen realizar posteriormente, como es el caso del análisis de Cluster que trabaja con índices de similitud, no obstante, estos se pueden transformar en índices de distancia y viceversa. A continuación se presentan los principales índices para los diferentes tipos de datos.

Tipo de dato

Binarios o de doble Estado

Índices

Emparejamiento simple (Sokal y Michener, 1958), Rogers y Tanimoto (1960), Hamman (1961), Jaccard (1908), Kulczynsky (1927), Russell y Rao (1940), Dice (1945), Ochiai (1957), Sokal y Sneath (1963), Yule (1911), Pearson.

Cualitativos

Extensión del emparejamiento simple, Distancia Chi-cuadrado (Benzecri, 1970).

Cuantitativos

Euclidean, Manhattan, Bray-Curtis, Canberra, Minkowski, Mahalanobis.

Datos mixtos

Gower (1971).

Genéticos

Nei (1972), Nei (1978), Hillis (1984), Swofford-Olsens (1990), Cavalli-Sforza y Edwards (1967) o distancia del arco, Cavalli-Sforza y Edwards (1965) o distancia de la cuerda, Rogers (1972), Prevosti (Wright, 1978).

Métodos multivariados. El origen del análisis multivariado se remonta a los comienzos del siglo XX, con Pearson y Sperman, época en la cual se empezaron a introducir los conceptos de la estadística moderna. Las bases definitivas de este tipo de análisis se establecieron en la década 1930-40 con Hotelling, Wilks, Fisher, Mahalanobis, y Bartlett (Bramardi, 2002). En términos generales, el análisis multivariado se refiere a todos aquellos métodos estadísticos que analizan simultáneamente medidas múltiples (más de dos variables) de cada individuo. En sentido estricto, son una extensión de los análisis univariados (análisis de distribución) y bivariados (clasificaciones cruzadas, correlación, análisis de varianza y regresiones simples) que se consideran como tal si todas las variables son aleatorias y están interrelacionadas (Hair et al., 1992).

En la caracterización de recursos fitogenéticos el análisis multivariado se puede definir como un conjunto de métodos de análisis de datos que tratan un gran número de mediciones sobre cada accesión del germoplasma. Su virtud principal consiste en permitir la descripción de las accesiones tomando en cuenta simultáneamente varias características, sin dejar de considerar la relación existente entre ellas. Los métodos multivariados se clasifican en dos grupos (Cuadro 5): (1) los de dependencia, que son aquellos en los cuales una variable o conjunto de variables es identificado como dependiente de otro conjunto conocidas como independiente o predictor; y (2) los de interdependencia, o aquellos en que ninguna variable o grupo de variables es definido como independiente o dependiente y, más bien, el procedimiento implica el análisis simultáneo de todo el conjunto de variables (Hair et al., 1992).

Cuadro 5. Clasificación de métodos estadísticos de análisis multivariados.

Métodos de dependencia (tipo de análisis)	Métodos de interdependencia (tipo de análisis)
Discriminante múltiple	Componentes principales
Correlación canónica	Factorial
Regresión múltiple	Conglomerados
Multivariante de la varianza	Multidimensional
Conjunto	Correspondencia

FUENTE: Hair et al., 1992.

Por su parte Bramardi (2002) puntualiza que para el caso del análisis de datos resultantes de caracterización de recursos genéticos vegetales (colecciones de germoplasma), el problema es representar geométricamente, cuantificar la asociación entre individuos y clasificarlos respecto a un conjunto de variables, las cuales pueden ser cuantitativas, cualitativas o la combinación de ambas. Teniendo en cuenta los objetivos que se desean alcanzar, este investigador clasifica los métodos multivariados en dos

grandes grupos. El primero se denomina de ordenación y permite arreglar y representar gráficamente el material en estudio en un número reducido de dimensiones. El segundo se denomina de clasificación y permite la búsqueda de grupos similares lo más homogéneos posible para clasificar los elementos en estudio de acuerdo con los métodos que aparecen en el Cuadro 6.

Cuadro 6. Clasificación de los métodos multivariados.

Métodos de ordenación	Métodos de clasificación
Análisis de componentes principales (ACP)	Análisis de cluster jerárquico –Arboles ultramétricos
Análisis canónico de poblaciones	Arboles aditivos
Análisis de coordenadas principales	Método de Ward (1963)
Análisis factorial de correspondencias	Clasificación no-jerárquica
Escalas multidimensionales	Arbol de mínima distancia
	Otros métodos

FUENTE: Adaptado de Bramardi, 2002.

A continuación se presenta un resumen de los métodos multivariados más usados para el análisis de datos de colecciones de germoplasma.

Análisis de componentes principales (ACP). Desde el punto analítico, este método se basa en la transformación de un conjunto de variables cuantitativas originales en otro conjunto de variables independientes no correlacionadas, llamadas componentes principales. Los componentes deben ser interpretados independientemente unos de otros, ya que contienen una parte de la varianza que no está expresada en otro componente principal (Pla, 1986; López e Hidalgo, 1994a).

Cuadro 7. Valores propios y proporción de la varianza explicada en análisis de componentes principales.

Componentes principales	Valores propios (λ_p)	Proporción de la varianza total explicada	
		Absoluta (%)	Acumulada (%)
Emergencia	4.535	30.225	30.235
Botón floral	3.157	21.045	51.280
Inicio floración	1.835	12.235	63.515
50% floración	0.962	6.415	69.930
Madurez Fisiológica	0.818	5.452	75.382
Índice cosecha	0.732	4.879	80.260
Rama principal	0.680	4.532	84.793
Dientes hojas	0.574	3.825	88.618
Diámetro tallo	0.498	3.318	91.936
Longitud panoja	0.339	2.260	94.196
Diámetro panoja	0.297	1.981	96.177
Altura planta	0.246	1.638	97.815
Diámetro grano	0.171	1.140	98.955
Peso 100 granos	0.099	0.658	99.613
Contenido de saponinas	0.058	0.387	100.00
Σ	15.00	100.00	–

FUENTE: Rojas, 1998.

El ACP concentra toda la variación presente en la matriz de datos originales en unos pocos ejes o componentes. Los componentes principales contienen información en diferentes proporciones de todas las variables originales y su número depende del número de éstas que se incorporen en el análisis. La contribución de las variables a cada componente principal se expresa en valores y vectores propios (ver ejemplo de valores propios en el Cuadro 7). El valor propio representa la varianza asociada con el componente principal y decrece a medida que se generan dichos componentes. En cambio, el vector propio contiene los coeficientes de las combinaciones lineales de las p variables originales.

El ACP es una herramienta útil para analizar los datos que se generan de la caracterización y evaluación preliminar de germoplasma, y permite conocer la relación existente entre las variables cuantitativas consideradas y la semejanza entre las accesiones; en el primer caso, con el fin de saber cuáles variables están o no asociadas, cuáles caracterizan en el mismo sentido o en el sentido contrario; y en el segundo, para saber cómo se distribuyen las accesiones, cuáles se parecen y cuáles no. También permite seleccionar las variables cuantitativas más discriminatorias para limitar el número de mediciones en caracterizaciones posteriores.

Una metodología alternativa que ayuda en la toma de decisiones relacionadas con los componente realmente importantes consiste en construir una gráfica de barras utilizando los valores de la varianza absoluta para cada componente en el eje Y, y los componentes principales en el eje X. Este método puede generar varias alternativas, algunas de las cuales son presentadas en la Figura 5. En el Caso 1, por ejemplo, el investigador puede seleccionar con un amplio margen de seguridad el primer componente principal (cp-1) que explica 71.2% de la varianza absoluta. En el Caso 2 puede seleccionar los cuatro primeros componentes principales (cp-1 a cp-4) que en conjunto explican 96.9% de la varianza. En el Caso 3 el ACP no parece apropiado y, por tanto, se debe buscar otra metodología que permita encontrar descriptores de mayor peso para explicar la variabilidad.

Análisis de conglomerados. Es un método analítico que se puede aplicar para clasificar las accesiones de un germoplasma (o variables) en grupos relativamente homogéneos con base en alguna similitud existente entre ellas. El objetivo en este análisis es clasificar un conjunto de n accesiones o p variables en un número pequeño de grupos o conglomerados, donde la formación de estos grupos puede obedecer a leyes naturales o a cualquier conjunto de características comunes a las accesiones.

El método de conglomerados o análisis de cluster se puede aplicar sobre una matriz básica de datos $n \times p$ o sobre una matriz $n \times n$, o $p \times p$, donde n es el número de accesiones que se quieren agrupar y p son las variables (Hair et al., 1992; López e Hidalgo, 1994b).

Es importante aclarar que el análisis de conglomerados se aplica sobre una matriz de distancias y no sobre una de similitud. Para descriptores cualitativos, esta última debe ser transformada en una de distancia. Para datos cuantitativos, los programas de estadística actualmente disponibles calculan directamente los valores de distancia según el método que se aplique.

Básicamente los métodos de agrupamiento más usados en el análisis conglomerado son: (1) jerárquico, que forma grupos a varios niveles; y (2) no-jerárquico o de partición que también forma grupos a través de criterios predefinidos (Recuadro 1) (Dillon y Goldstein, 1984; López e Hidalgo, 1994b).

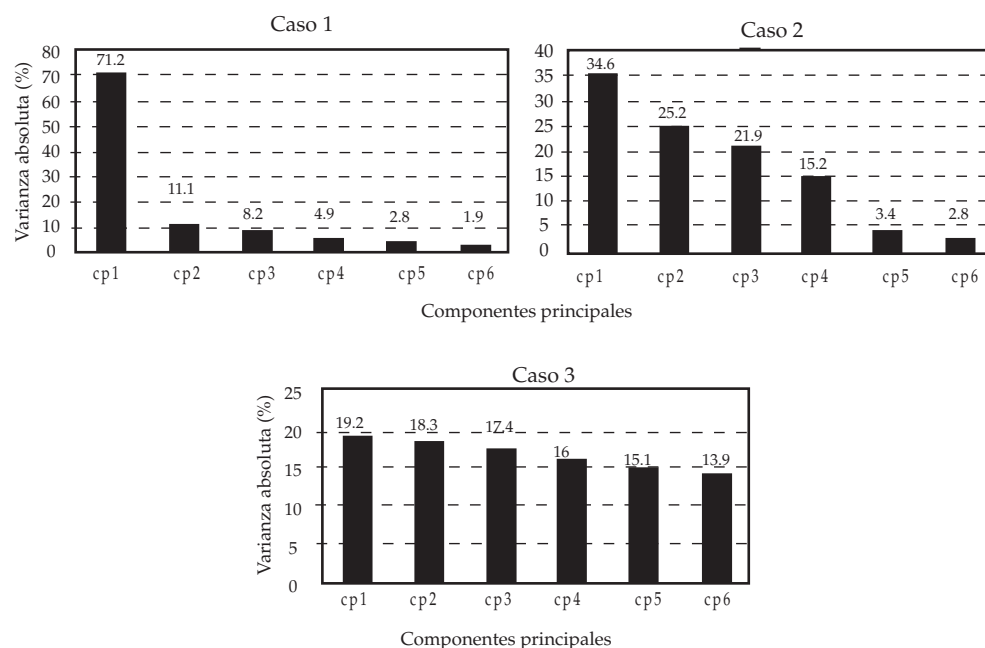


Figura 5. Ejemplo de comparaciones de las varianzas absolutas de seis componentes principales en los casos presentados en este documento.

Recuadro 1. Diferencias entre los métodos de aglomeración.

Método jerárquico	Método no - jerárquico
1. El agrupamiento parte de matriz simétrica ($n \times n$).	1. El agrupamiento parte de matriz básica de datos ($n \times p$).
2. Usa menor volumen de datos.	2. Usa mayor volumen de datos.
3. No da margen para la reasignación de objetos a los grupos.	3. La asignación de un objeto a un grupo puede ser cambiada.
4. Usa dendrogramas.	4. No usa dendrogramas.
5. Los datos son muy susceptibles a contener observaciones extrañas.	5. Los datos son menos susceptibles a contener observaciones extrañas.

FUENTE: Dillon y Goldstein (1984), Hair et al. (1992).

Agrupamiento jerárquico. Se caracteriza por sucesivas fusiones para formar los grupos. Algunos de estos grupos tienen mayor rango y cada uno de ellos abarca varios de menor orden permitiendo, de esta manera, seguir en detalle la formación de los conglomerados y conocer el nivel de similitud al que se agrupa cada conjunto de individuos (Dillon y Goldstein, 1984).

Por las fusiones sucesivas que realiza también se le conoce como método jerárquico aglomerativo. El procedimiento parte de la existencia inicial de un conglomerado para cada accesión o variable que por aproximaciones sucesivas se van uniendo con otras en grupos hasta formar un conglomerado único, que los incluye a todas. En el agrupamiento los resultados se presentan en forma de diagrama de árbol, más comúnmente conocido como dendrograma (López e Hidalgo, 1994b).

Para la formación de los conglomerados por este método jerárquico existen diversas formas de enlace, siendo las más comunes: Simple, completo, UPGMA, centroide y

'Ward', todas con el mismo criterio de maximizar la variación entre los grupos y minimizarla dentro de ellos (Dillon y Goldstein, 1984; Hair et al., 1992).

Agrupamiento no-jerárquico. Esta forma también es conocida como de partición. Se caracteriza por dividir el grupo de accesiones en un número preseleccionado de conglomerados que no tienen una estructura jerárquica. Desde el punto de la estadística, se caracteriza por maximizar la suma de cuadrados intersujetos entre conglomerados con respecto a la suma de cuadrados intrasujetos (López e Hidalgo, 1994b).

Este método tiene muchas variantes, siendo la más utilizada 'k-means' que se caracteriza por centrar de nuevo cada accesión cuando cambia de grupo. El cálculo se inicia a partir de una matriz rectangular o matriz básica de datos ($n \times p$), lo que hace posible mantener un mayor volumen de estos que en los agrupamientos jerárquicos. Además, este agrupamiento tiene como ventaja permitir la redistribución de una accesión de un conglomerado a otro cuando su asignación inicial es incorrecta, como resultado de las sucesivas lecturas que se realizan sobre los mismos datos (López e Hidalgo, 1994a; Crivisqui, 1998).

En el Recuadro 1 también se pueden observar las diferencias más importantes entre ambos métodos de agrupamiento. Hair et al. (1992) consideran que no es posible dar una respuesta definitiva sobre cuál método se debe utilizar. La investigación del problema en forma típicamente manual posiblemente sugiere un método e, igualmente, mediante un ordenador es posible desarrollar rápidamente ambos métodos y su orden si son para aplicaciones futuras. En los estudios de caso sobre jícama (Caso 3) y pasifloras andinas (Caso 4) que aparecen más adelante en este boletín se muestra la aplicación de estos métodos.

Cuadro 8. Métodos multivariados de acuerdo con el tipo de variables e índices de similitud y distancias asociadas.

Tipo de método	Tipo de variable	Índice o distancia asociada	Método multivariado
Ordenación	Cuantitativas	Euclídea Mahalanobis	Componentes principales. Análisis canónico de poblaciones.
	Cualitativas	Similitud Datos de frecuencia organizados en tablas de contingencia. Distancia Chi-cuadrado	Análisis de coordenadas principales. Análisis factorial de correspondencias.
	Cuantitativas o cualitativas	Índice de disimilitud	Análisis de proximidades (MDS o escalas multidimensionales).
Clasificación	Cuantitativas o cualitativas	Según tipo de variables, asociar una matriz de distancia o similitud adecuada	Análisis Cluster.

FUENTE: Adaptado de Bramardi, 2002.

Los métodos no-jerárquicos tienen una gran aceptación y sus aplicaciones se están incrementando, no obstante su uso depende de la habilidad del investigador para seleccionar centros móviles o individuos típicos a partir de los cuales se inicia el proceso de aglomeración, lo que depende de la práctica, los objetivos y la base teórica. En este sentido estos métodos tienen varias ventajas sobre los métodos jerárquicos (Hair et al., 1992).

Dependiendo del tipo de variables que se hayan registrado del material en estudio, Bramardi (2002) propone la guía que aparece en el Cuadro 8 y que facilita la presentación y selección de los métodos multivariados y sus coeficientes de similitud o de distancia asociados.

En la sección siguiente se incluyen cuatro estudios de caso que tienen como objetivo conocer la variabilidad genética de especies cultivadas. En estos casos se pueden ver los diferentes enfoques en la caracterización y los métodos utilizados para el análisis de los datos en cada especie.

Referencias

- Bramardi, S. J. 2002. Análisis multivariado. Su aplicación en la caracterización de recursos genéticos. Facultad de Ciencias Agrarias, Univ. Conahue, Estación Exp. INTA, Argentina. 60 p. (manuscrito).
- Crisci, J. V. y López, M. F. 1983. Introducción a la teoría y práctica de la taxonomía numérica. Secretaría General de la Organización de los Estados Americanos (OEA). Washington, D.C. 132 p.
- Crivisqui, E. 1998. Presentación de los métodos de clasificación. En: Seminario de métodos estadísticos multivariados. Valdivia, Chile. 5-16 de enero de 1998. 56 p.
- Cupolillo, E.; Momen, H.; y Grimaldi, G. Jr. 1998. Genetic diversity in natural population of new world *Leishmania*. Men. Inst. Oswaldo Cruz 93(5):663-668.
- Dillon, W. R. y Goldsten, M. 1984. Multivariate analysis: Methods and applications. Wiley, Nueva York. 587 p.
- Enríquez, G. 1966. Selección y estudio de las características de la flor, la hoja y la mazorca, útiles para la identificación y descripción de cultivares de cacao. Tesis. MSc. Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Turrialba, Costa Rica. 97 p.
- Enríquez, G. 1991. Descripción y evaluación de los recursos genéticos. En: Castillo, R. y Estrella, J. Tapia, C. (eds.). Técnicas para el manejo y uso de recursos genéticos vegetales. Editorial Porvenir. Quito, Ecuador. p. 116 160.
- FAO. 1996. Informe sobre el estado de los recursos fitogenéticos en el mundo. En: Conferencia Técnica Internacional sobre los Recursos Fitogenéticos. Leipzig, Alemania, junio 17-23. 1996. FAO, Roma. p. 75.
- Ford-Lloyd, B. y Jackson, M. 1986. Plant genetic resources. An introduction to their conservation and use. Edward Arnold, Publ. Baltimore. 152. p.
- Hawkes, J. G. 1970. The taxonomy of cultivated plants. En: O. H. Frankel and E. Bennet (eds.). Genetic resources in plants. Their exploration and conservation. IBP Handbook no. 11.
- Hutchinson, J. 1964. The genera of flowering plants (Angiospermae). Vol. 1. Dicotyledonea. Oxford Univ. Press, Londres.
-

- Hair, J. F.; Anderson, R. E., Tatham, R. L.; y Black, W. C. 1992. Multivariate data analysis. MacMillan Publ. Co. Nueva York. 544 p.
- López J. A. e Hidalgo, M. D. 1994a. Análisis de componentes principales y análisis factorial. En: Ato, M. y López, J. J. (eds.). Fundamentos de estadística con Systat. Addison Wesley Iberoamericana. p. 457-503.
- López J. A. e Hidalgo, M. D. 1994b. Análisis de conglomerados. En: Ato, M. y López, J. J. (eds.). Fundamentos de estadística con Systat. Addison Wesley Iberoamericana. p. 505-532.
- Pla, L. E. 1986. Análisis multivariado: Método de componentes principales. Secretaría de la Organización de Estados Americanos (OEA). Washington, D.C. 94 p.
- Plucknett, D. L.; Williams, J. T.; Smith, N.; y Anishetty, N. N. 1992. Los bancos genéticos y la alimentación mundial. Editorial IICA, San José, Costa Rica. 257 p.
- Plucknett, D. L. et. al. 1992. Recolectores de plantas y bancos genéticos. En: los bancos genéticos y la alimentación mundial. Instituto Interamericano de Cooperación para la Agricultura (IICA), San José, Costa Rica. p. 57-86.
- Pound, J.F. 1932. The genetic constitution of the cacao crop. En: Imperial College of Tropical Agriculture, Trinidad. Annual Report on Cacao Research 2:27-29.
- Rojas, W. 1998. Análisis de la diversidad genética del germoplasma de quinua (*Chenopodium quinoa* Willd.) de Bolivia, mediante métodos multivariados. Tesis MSc. Escuela de Graduados, Facultad de Ciencias Agrarias, Universidad Austral de Chile. 209 p.
- Steel, G. y Torrie, J. 1980. Principles and procedures of statistics: A biometrics approach. McGraw-Hill, Nueva York. (s.p.).
- Watson, J. D. y Crick, C. 1953. Molecular structure of nucleic acids. En: Classic paper in genetic. Prentice Hall. p. 241-243.
-

Caracterización Morfológica de Germoplasma Estudios de casos

Caso 1. Análisis de la variabilidad genética en quinua

Wilfredo Rojas*

En este caso se aplicarán los métodos multivariados descritos en la primera parte de este boletín para analizar en forma parcial la información generada en la caracterización y evaluación preliminar del germoplasma de quinua (*Chenopodium quinoa* Willd.), incluyendo 15 variables cuantitativas sobre 1512 accesiones cultivadas en las temporadas agrícolas 1992-93 y 1993-94.

El objetivo es estudiar la variabilidad genética asociada con estas variables con el fin de saber cómo se estructuran, cuáles están asociadas y cuáles caracterizan en un mismo sentido o en sentido contrario.

Se consideran las variables cuantitativas discontinuas: emergencia de plántulas, botón floral, inicio de floración, 50% de floración ($F_{(50\%)}$), madurez fisiológica, número de ramas y número de dientes; y las variables cuantitativas continuas: diámetro de tallo, longitud y diámetro de panoja, altura de planta, diámetro de grano, peso de 100 granos, contenido de saponinas e índice de cosecha. Con el fin de alcanzar el objetivo del estudio, el análisis de las accesiones se efectuó mediante: (1) correlación simple, (2) componentes principales, y (3) conglomerados a través de la técnica de agrupamiento jerárquico.

Coeficiente de correlación simple. Conocido como coeficiente de Pearson, sirve para medir en términos relativos el grado de asociación entre pares de características. Se recomienda cuando las unidades de medida de las variables son diferentes, por ejemplo, aparición del botón floral (días), diámetro de tallo (mm), altura de planta (cm), peso de 100 granos (g) y contenido de saponinas (cs).

Componentes principales. Este método realiza una transformación lineal sobre las variables originales y permite generar un nuevo conjunto de variables independientes o componentes principales.

Conglomerados. Para construir el agrupamiento o dendrograma se aplicaron la distancia euclidiana al cuadrado y el algoritmo de enlace completo que se caracteriza por usar una regla de distancia máxima que comienza evaluando las dos variables más distantes entre sí para formar el primer conglomerado. En una segunda etapa el algoritmo busca una nueva variable para aumentar el número de elementos del conglomerado formado y, así sucesivamente, hasta que todas las variables conformen un solo conglomerado (Hair et al., 1992; López e Hidalgo, 1994b).

* Especialista en Recursos Fitogenéticos, Fundación PROINPA, Regional Altiplano, La Paz, Bolivia.

Presentación y discusión de los resultados

Para facilitar la interpretación de los resultados, las variables fueron agrupadas de la manera siguiente: (1) Fenológicas: emergencia de plántulas, botón floral, inicio de floración, 50% de floración y madurez fisiológica; (2) de arquitectura de planta: número de ramas, número de dientes en las hojas, diámetro de tallo, longitud y diámetro de panoja y altura de planta; (3) de grano: diámetro de grano, peso de 100 granos y contenido de saponina; y (4) índice de cosecha.

Análisis de correlación simple. La matriz de correlación entre cada par de características cuantitativas se presenta en el Cuadro 1.1 donde se observa que 83 coeficientes fueron altamente significativos ($P \leq 0.001$). Se consideró que los coeficientes > 0.40 corresponden a asociaciones que representan patrones naturales de variación; de esta forma, las correlaciones más importantes fueron las variables fenológicas y las de grano, en comparación con las variables de arquitectura de planta.

Entre las variables fenológicas, la correlación más alta correspondió al inicio de floración y 50% de floración ($r = 0.94$). Estas variables, en su orden, están altamente correlacionadas con la aparición del botón floral ($r = 0.69$ y $r = 0.73$); madurez fisiológica ($r = 0.63$ y $r = 0.61$) e índice de cosecha ($r = -0.59$ y $r = -0.57$). Las correlaciones negativas con el índice de cosecha indican que esta característica tiende a ser menor a medida que las fases fenológicas son más tardías, lo cual es corroborado por las correlaciones negativas con la madurez fisiológica ($r = -0.55$) y la aparición del botón floral ($r = -0.42$).

La correlación positiva de la madurez fisiológica con altura de planta ($r = 0.56$) y diámetro de tallo ($r = 0.41$) indica que las plantas tienden a un mayor desarrollo de estas características cuando la duración del ciclo fenológico es más tardío; sin embargo, la correlación negativa de la madurez con el peso de 100 granos ($r = -0.42$) indica que también existe la tendencia hacia bajos índices de cosecha.

Las asociaciones positivas entre el 50% de floración y el número de ramas ($r = 0.44$) y dientes en la hoja ($r = 0.45$), y entre el inicio de floración y el número de ramas ($r = 0.43$) indican que las accesiones que más tardan en florecer se caracterizan por desarrollar una mayor ramificación y hojas más dentadas, lo cual fue evidente en las accesiones procedentes de los valles.

Otra asociación importante fue la formada por el diámetro de grano y el peso de 100 granos ($r = 0.89$), a su vez, el diámetro de grano se correlacionó positivamente con el contenido de saponinas ($r = 0.40$) y negativamente con la longitud de panoja ($r = -0.40$), indicando que las accesiones con granos grandes tienden a desarrollar alto contenido de saponinas y panojas de corta longitud.

Entre las variables de arquitectura de planta, el diámetro del tallo presentó una correlación positiva con la altura de planta ($r = 0.69$), con el diámetro de panoja ($r = 0.60$) y con la longitud de panoja ($r = 0.40$), y esta última, a su vez, con la altura de planta ($r = 0.58$). Lo anterior significa que las accesiones que durante las primeras fases fenológicas desarrollan mayor diámetro de tallo y altura de planta también desarrollan un mayor tamaño de panoja.

Análisis de componentes principales (ACP). En este estudio, los resultados de los componentes principales son interpretados tomando como base sus valores y vectores propios. Los valores propios y la varianza total explicada por cada uno de los componentes, así como la proporción de la varianza total, se incluyen en el Cuadro 1.2.

Cuadro 1.1. Ejemplo de una matriz de correlación simple entre 15 variables cuantitativas utilizadas para caracterizar germoplasma de quinua (n = 1512).

Características	EM	BF	IF	F _(50%)	MF	IC	RP	DH	DT	LP	DP	AP	DG	P100	CS
Emergencia	1.00														
Botón floral	0.14	1.00													
Inicio floración	0.21	0.69	1.00												
50% floración	0.16	0.73	0.94	1.00											
Madurez Fisiológica	0.25	0.37	0.63	0.61	1.00										
Indice cosecha	-0.09	-0.42	-0.59	-0.57	-0.55	1.00									
Rama principal	-0.01	0.37	0.39	0.44	0.16	-0.33	1.00								
Dientes hojas	0.05	0.38	0.43	0.45	0.23	-0.31	0.25	1.00							
Diámetro tallo	0.22	0.11	0.28	0.28	0.41	-0.20	0.31	0.12	1.00						
Longitud panoja	0.19	-0.30	-0.09	-0.15	0.34	0.02	-0.23	-0.24	0.40	1.00					
Diámetro panoja	0.12	-0.06	0.04	0.06	0.20	-0.02	0.20	0.01	0.60	0.31	1.00				
Altura planta	0.08	0.08	0.30	0.30	0.56	-0.32	0.27	0.10	0.69	0.58	0.38	1.00			
Diámetro grano	-0.21	0.03	-0.09	-0.07	-0.38	0.10	0.23	0.15	-0.17	-0.40	-0.05	-0.24	1.00		
Peso 100 granos	-0.22	-0.01	-0.16	-0.14	-0.42	0.17	0.14	0.15	-0.21	-0.36	-0.08	-0.25	0.89	1.00	
Cont. saponinas	-0.11	-0.13	0.12	0.16	-0.01	-0.11	0.21	0.23	0.01	-0.28	0.04	-0.08	0.40	0.34	1.00
	EM	BF	IF	F _(50%)	MF	IC	RP	DH	DT	LP	DP	AP	DG	P100	CS

Valores en negrilla no son significativos (P ≥ 0.001).
 FUENTE: Rojas, 1998.

Cuadro 1.2. Valores propios y proporción de la varianza explicada en el análisis de los componentes principales en la caracterización de una colección de quinua.

Componentes principales	Valores propios (λ_p)	Proporción de la varianza total explicada	
		Absoluta (%)	Acumulada (%)
Emergencia (días)	4.535	30.225	30.235
Botón floral (días)	3.157	21.045	51.280
Inicio de floración (días)	1.835	12.235	63.515
50% de floración (días)	0.962	6.415	69.930
Madurez fisiológica (días)	0.818	5.452	75.382
Índice de cosecha	0.732	4.879	80.260
Ramas principales (no.)	0.680	4.532	84.793
Dientes en las hojas (no.)	0.574	3.825	88.618
Diámetro de tallo (mm)	0.498	3.318	91.936
Longitud de panoja (cm)	0.339	2.260	94.196
Diámetro de panoja (cm)	0.297	1.981	96.177
Altura de planta (cm)	0.246	1.638	97.815
Diámetro del grano (mm)	0.171	1.140	98.955
Peso de 100 granos (g)	0.099	0.658	99.613
Contenido de saponina (cs)	0.058	0.387	100.00
Σ	15.00	100.00	—

FUENTE: Rojas, 1998.

Se observa que la varianza asociada con cada componente principal es diferente y decrece en orden. El primer componente explica el 30.23% de la varianza total, el segundo explica el 21.04%, y así sucesivamente, hasta que toda la variabilidad queda distribuida diferencialmente entre los 15 componentes.

La selección del número de componentes que se debe retomar para el análisis es aún tema de discusión entre especialistas, ya que no existen pruebas estadísticas inferenciales que permitan probar la significancia de dichos valores; sin embargo, debe quedar claro que la selección no depende del número de componentes obtenido ya que el análisis genera tantos componentes como variables hay en el estudio (Dillon y Goldstein, 1984; López e Hidalgo, 1994a). En general existen diversos criterios de selección que varían de acuerdo con las decisiones del investigador y que ayudan a tomar tal decisión, algunos de ellos se citan a continuación.

- Cliff en 1987 indicó que se deben considerar como aceptables los componentes cuyos valores propios expliquen un 70% o más de la varianza total (López e Hidalgo, 1994a).
- Kaiser en 1960 estableció un criterio utilizado frecuentemente y que consiste en la selección de los componentes cuyo valor propio sea ≥ 1 (López e Hidalgo, 1994a).
- Catell en 1966 sugirió un criterio gráfico que consiste en representar el número de componentes y su valor propio en la abscisa, y el porcentaje de la varianza correspondiente en la ordenada, lo que permite observar en forma gráfica el decrecimiento de los primeros componentes en relación con los demás (Pla, 1986).

Teniendo en cuenta el criterio de Cliff en la interpretación y toma de decisiones de los datos presentados en el Cuadro 1.2 se tendrían que seleccionar los primeros cinco

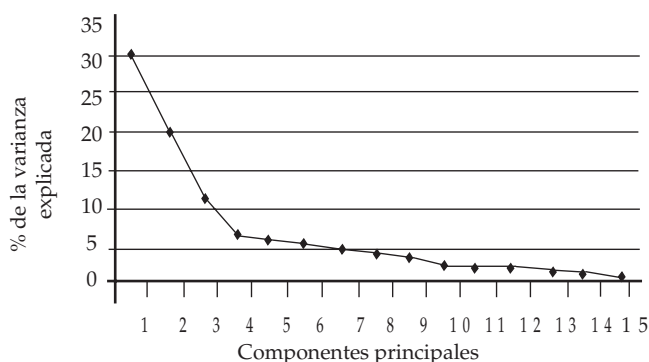


Figura 1.1. Proporción de la varianza explicada por cada componente principal en la caracterización de una colección de quinua.

componentes ya que en conjunto explican más del 75% de la varianza, según el criterio gráfico se deben retener los cuatro primeros (Figura 1.1) que en conjunto explican más del 69% de la varianza y según Kaiser se deberían seleccionar los tres primeros que en conjunto explican más del 63% de la variación total.

De acuerdo con los criterios anteriores, para la toma de decisiones sobre el número de componentes se debe recurrir a los vectores propios (Cuadro 1.3) y a la correlación entre las variables originales y los cinco componentes principales (Cuadro 1.4), así como al análisis detallado del conjunto de coeficientes que se obtienen de las variables originales asociadas con cada componente. Este proceso se debe continuar hasta obtener una información suficiente, ya que la inclusión de cada componente implica también su descripción.

La interpretación en los vectores propios (Cuadro 1.3) y la correlación entre las variables originales y los componentes principales se deben centrar en los coeficientes; mientras más altos sean estos, independientemente del signo, más eficientes serán en la discriminación de las accesiones. Las variables con coeficiente negativo (-) significan que están caracterizando en sentido contrario en relación con las variables positivas (+) y viceversa. Este último aspecto fue confirmado por Ferreira (1987) quien sugirió que las cargas que se distribuyen en los componentes indican el peso de cada variable asociada o el grado de contribución al componente, por tanto, recomienda tomar en cuenta el comportamiento observado en las accesiones durante el trabajo de caracterización en relación con cada variable considerada en el estudio.

Por lo general los programas estadísticos proporcionan la información de los vectores propios y en algunos casos la correlación entre las variables originales y los componentes principales. Sin embargo, los valores en el Cuadro 1.4 se pueden calcular a partir de los vectores propios (Cuadro 1.3) mediante la fórmula siguiente (Pla, 1986).

$$r_{(jk)} = l_{(jk)} \times (\lambda_{(k)})^{1/2} / s_{(jj)}$$

donde,

$r_{(jk)}$ = correlación entre la variable $x_{(j)}$ original y el k -ésimo componente.

$l_{(jk)}$ = elemento j -ésimo del k -ésimo vector propio.

$\lambda_{(k)}$ = k -ésimo valor propio.

$s_{(jj)}$ = varianzas de la matriz de correlación.

Cuadro 1.3. Vectores propios de los primeros cinco componentes principales en la caracterización de una colección de quinua.

Características	Componentes principales				
	1	2	3	4	5
Emergencia (días)	0.138	-0.128	-0.091	0.873	0.121
Botón floral (días)	0.296	0.262	-0.188	0.123	-0.195
Inicio de floración (días)	0.400	0.168	-0.154	0.029	-0.005
50% de floración (días)	0.399	0.195	-0.136	0.002	-0.057
Madurez fisiológica (días)	0.382	-0.126	-0.084	-0.105	0.270
Índice de cosecha	-0.322	-0.099	0.104	0.255	-0.143
Ramas principales (no.)	0.221	0.221	0.260	-0.080	-0.558
Dientes en las hojas (no.)	0.206	0.254	0.013	0.086	0.336
Diámetro de tallo (mm)	0.084	-0.195	0.413	0.100	-0.111
Longitud de panoja (cm)	0.131	-0.442	0.176	-0.068	0.311
Diámetro de panoja (cm)	0.293	-0.173	0.478	0.200	-0.184
Altura de planta (cm)	0.293	-0.235	0.318	-0.251	0.078
Diámetro del grano (mm)	-0.132	0.400	0.360	0.073	0.091
Peso de 100 granos (g)	-0.164	0.378	0.341	0.077	0.145
Contenido de saponina (cs)	0.031	0.297	0.236	-0.038	0.495
Σ^2	1.000	1.000	1.000	1.000	1.000

FUENTE: Rojas, 1998.

Cuadro 1.4. Correlación entre las variables originales y los cinco primeros componentes principales en la caracterización de una colección de quinua.

Características	Componentes principales				
	1	2	3	4	5
Emergencia (días)	0.294	-0.228	-0.123	0.857	0.110
Botón floral (días)	0.631	0.466	-0.255	0.121	-0.17
Inicio de floración (días)	0.851	0.298	-0.208	0.029	-0.005
50% de floración (días)	0.849	0.346	-0.184	0.002	-0.052
Madurez fisiológica (días)	0.813	-0.223	-0.114	-0.103	0.245
Índice de cosecha	-0.686	-0.176	0.141	0.251	-0.130
Ramas principales (no.)	0.471	0.393	0.352	0.076	0.13
Dientes en las hojas (no.)	0.438	0.451	0.017	-0.079	-0.505
Diámetro de tallo (mm)	0.588	-0.346	0.560	0.085	0.304
Longitud de panoja (cm)	0.178	-0.785	0.238	0.099	-0.101
Diámetro de panoja (cm)	0.279	-0.308	0.648	-0.067	0.282
Altura de planta (cm)	0.623	-0.418	0.420	0.197	-0.167
Diámetro del grano (mm)	-0.282	0.711	0.488	-0.247	0.071
Peso de 100 granos (g)	-0.350	0.672	0.462	0.072	0.083
Contenido de saponina (cc)	0.065	0.527	0.320	-0.038	0.448
Σ^2 = valor propio	4.535	3.157	1.835	0.962	0.818

FUENTE: Rojas, 1998.

Considerando que las varianzas $s_{(ij)}$ son unitarias en la matriz de correlación (Cuadro 1.1), el cálculo para la primera variable (emergencia) del primer componente principal utilizando los datos de los Cuadros 1.2 y 1.3, es el siguiente:

$$r = 0.138 \times (4.535)^{1/2} = 0.294$$

La correlación ($r = 0.294$) representa la asociación entre la variable emergencia de

plántulas con el primer componente principal (Cuadro 1.4). Este mismo procedimiento se debe utilizar para calcular las correlaciones entre las demás variables y los componentes. Es necesario notar que en este caso la suma al cuadrado de los coeficientes de correlación del primer componente (4.535) es igual al primer valor propio en el Cuadro 1.2.

Para la selección de los valores propios significativos se escogió el criterio de Kaiser ya que se adapta bien a la finalidad del presente análisis y, por consiguiente, la descripción de los resultados se hará en función de los tres primeros componentes que explican más del 63% de la varianza total (ver Cuadro 1.2). Estos componentes, a su vez, dan una idea clara de la estructura que subyace a las variables cuantitativas sometidas al análisis.

La interpretación de los resultados de los valores propios (Cuadro 1.2), vectores propios (Cuadro 1.3) y correlación de las variables originales (Cuadro 1.4) sobre los tres primeros componentes seleccionados se presenta a continuación.

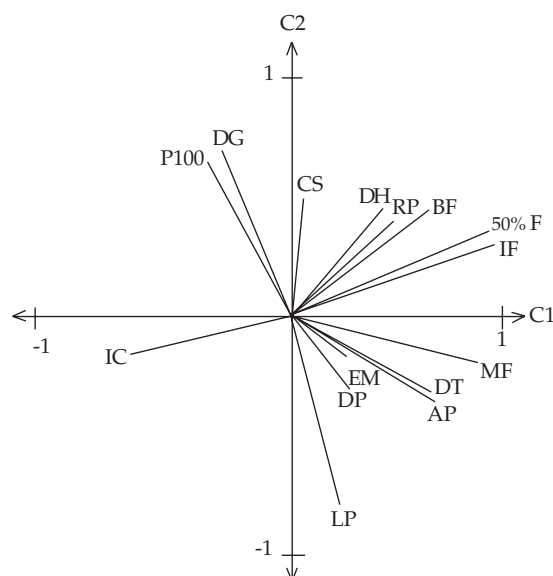
El primer componente principal contribuyó con más del 30% de la varianza total explicada, mientras que la distribución de los coeficientes del primer vector propio y de correlación indica que el inicio de floración, 50% de floración y la madurez fisiológica fueron las variables que más contribuyeron en forma positiva a dicho componente. De forma secundaria lo hicieron el botón floral, la altura de planta y el diámetro de tallo; por el contrario, el índice de cosecha, el diámetro de grano y el peso de 100 granos fueron las variables que más contribuyeron en forma negativa.

Los resultados anteriores indican que el primer componente permitió distinguir las accesiones que florecen y maduran en forma tardía y que registran igualmente valores bajos en índices de cosecha. De la misma manera, la contribución positiva de la altura de planta y el diámetro del tallo indican que la quinua, además de tardía, desarrolla plantas con arquitectura prominente que incide negativamente en el índice de cosecha, aunque los signos negativos del diámetro de grano y el peso de 100 granos sugieren que los índices bajos también se deben en parte a la formación de granos pequeños.

El número de dientes en las hojas y de ramas primarias también contribuyeron en forma positiva a este primer componente, aunque en menor proporción que la altura de planta y el diámetro de tallo, lo cual significa que las accesiones tardías también tienden a formar más dientes en sus hojas y a presentar una mayor ramificación.

El segundo componente principal contribuyó con más del 21% de la varianza total explicada (Cuadro 1.2). De acuerdo con los coeficientes del segundo vector propio (Cuadro 1.3) y de correlación (Cuadro 1.4) las variables que más contribuyeron en forma positiva fueron el diámetro de grano, el peso de 100 granos y el contenido de saponinas; por el contrario, la mayoría de las demás variables contribuyeron en forma negativa destacándose entre ellas la longitud de la panoja. En conclusión, con este segundo componente fue posible distinguir las accesiones de quinua que forman los granos más grandes con altos contenidos de saponinas y que desarrollan, igualmente, plantas ligeramente ramificadas con hojas relativamente dentadas y tamaño de panoja, altura y diámetro de tallo pequeños.

El tercer componente principal contribuyó con más del 12% de la varianza total explicada (Cuadro 1.2). En este caso, los coeficientes del tercer vector propio y de correlación indican que los diámetros de panoja y de tallo fueron las características que más contribuyeron, seguidas del diámetro de grano, peso de 100 granos y altura de planta. En consecuencia, este componente permitió distinguir las accesiones con plantas de mayor



P100 = Peso de 100 granos. DG = Diámetro de grano. CS = Contenido de saponina. DH = Diámetro de hoja. RP = Rendimiento por planta. BF = Botón floral. 50%F = 50% de floración. IF = Inicio de floración. MF = Madurez fisiológica. DT = Diámetro de tallo. AP = Altura de planta. EM = Emergencia de planta. DP = Diámetro de panoja. LP = Longitud de panoja. IC = Índice de cosecha.

Figura 1.2. Distribución de las variables originales de accesiones sobre el primero y segundo componente principal en la caracterización de una colección de quinua.

FUENTE: Rojas, 1998.

altura, tallos gruesos, panojas grandes y con granos desde tamaños medianos hasta grandes.

Por tratarse en este caso de un análisis normado (con datos estandarizados), las coordenadas de las variables sobre cada componente principal son iguales a la correlación entre las variables originales y los componentes principales (Cuadro 1.4); por tanto, es posible representar la proyección de las variables originales sobre los dos primeros ejes principales (Figura 1.2). En el caso del índice de cosecha (IC), el procedimiento para ubicar las variables sobre el espacio bidimensional consiste en proyectar -0.686 en el primer eje y -0.176 en el segundo eje, y así sucesivamente, hasta completar el círculo de correlaciones.

En la Figura 1.2 también se observa que las variables más vinculadas en forma positiva con el primer eje son el inicio de floración (IF), 50% de floración ($F_{(50\%)}$) y madurez fisiológica (MF); y en forma negativa el índice de cosecha (IC). Las variables más vinculadas al segundo eje en sentido positivo son peso de 100 granos (P100), diámetro de grano (DG) y contenido de saponinas (CS); y en sentido negativo la longitud de panoja (LP).

La proyección opuesta del índice de cosecha sobre el primer eje en relación con las variables fenológicas significa que las accesiones de quinua desarrollan un menor valor de este índice en la medida que tardan en florecer y en alcanzar la madurez; o también, a medida que son más precoces desarrollan un mayor índice de cosecha.

Un procedimiento de interpretación similar se recomienda seguir para la proyección de las variables sobre el segundo eje, con el fin de comprobar si las accesiones de quinua tienden a desarrollar mayor tamaño de grano y contenido de saponinas en la medida que la longitud de panoja sea menor, y viceversa.

En la Figura 1.2 también se puede observar el grado de asociación entre las variables que está determinada por la separación angular que forman sus proyecciones. Se recomienda, además, considerar las distancias entre cada una de éstas a partir del origen, siendo su contribución mayor mientras más distantes se encuentren. De acuerdo con la separación angular, la mejor asociación está constituida por altura de planta (AP) con el diámetro de tallo (DT) y emergencia de plántula (EM), seguida por inicio de floración (IF) con 50% de floración ($F_{(50\%)}$) y peso de 100 granos (P100) con el diámetro de grano (DG). La distancia al origen indica que las variables florales y de grano son las más importantes.

Proporción de la varianza explicada. En estudios de esta naturaleza es importante determinar el grado de discriminación de las variables en estudio con el objeto de identificar tanto las de mayor como las de menor variación dentro del germoplasma.

A través del análisis de los componentes principales es posible determinar el grado de discriminación, cuantificando la proporción de varianza explicada por cada variable original sobre los tres componentes seleccionados; para ello, es necesario efectuar la suma al cuadrado de la correlación que forma cada variable original con los tres componentes (ver Cuadro 1.4). Esta operación es factible debido a que los componentes no están correlacionados entre sí (Pla, 1986). Para el caso de la variable 50% de floración sería: $(0.849)^2 + (0.346)^2 + (-0.184)^2 = 0.875$, siendo similar el cálculo con las demás variables. En el Cuadro 1.5 se presentan los resultados con las variables ordenadas en forma descendente. Se debe tener en cuenta que las variables que explican una mayor

Cuadro 1.5. Proporción de la varianza explicada por cada variable original sobre los tres primeros componentes principales en la caracterización de una colección de quinua.

Características	Componentes principales			Proporción de la varianza
	1	2	3	
50% de floración (días)	0.721	0.120	0.034	0.875
Inicio de floración (días)	0.724	0.089	0.043	0.856
Diámetro del grano (mm)	0.080	0.506	0.238	0.824
Peso de 100 granos (g)	0.123	0.452	0.213	0.788
Diámetro de tallo (mm)	0.346	0.120	0.314	0.780
Altura de planta (cm)	0.388	0.175	0.176	0.739
Madurez fisiológica (días)	0.661	0.050	0.013	0.724
Longitud de panoja (cm)	0.032	0.616	0.057	0.705
Botón floral (días)	0.398	0.217	0.065	0.680
Diámetro de panoja (cm)	0.078	0.095	0.420	0.593
Índice de cosecha	0.471	0.031	0.020	0.522
Ramas principales (no.)	0.222	0.154	0.124	0.500
Dientes en las hojas (no.)	0.192	0.203	0.001	0.396
Contenido de saponinas (cs)	0.004	0.278	0.102	0.384
Emergencia (días)	0.086	0.052	0.015	0.154

proporción de varianza son las más discriminatorias y, por tanto, su importancia es mayor.

En general, los resultados en el Cuadro 1.5 muestran que en las accesiones de quinua las variables fenológicas son más importantes y discriminatorias que las variables de grano y éstas, a su vez, más que las variables que describen la arquitectura de la planta. De otro modo, se puede interpretar que las variables reproductivas son más discriminatorias que las vegetativas.

Entre las variables fenológicas, el 50% de floración y el inicio de floración ocupan el primero y segundo lugar, respectivamente, por tanto, son las más discriminatorias del germoplasma, cuyo comportamiento depende principalmente del genotipo; la madurez fisiológica –en el séptimo lugar– tiene una posición aceptable para estudiar la variabilidad genética; mientras que la emergencia de plántulas en la última posición es de menor variación en el germoplasma.

El grado de discriminación de la variable emergencia de plántulas permite deducir que su comportamiento no depende tanto del genotipo sino, más bien, de otros factores como la preparación del suelo y la humedad en el momento de la siembra, por tanto, no se recomienda su inclusión en futuros estudios con este cultivo.

El diámetro de grano y el peso de 100 granos –en tercero y cuarto lugar, respectivamente– también son importantes y su comportamiento depende del genotipo. Es necesario señalar que el contenido de saponinas en la planta aparece en el penúltimo lugar, no obstante que en el germoplasma existen accesiones desde dulces (sin saponinas) hasta muy amargas (con alto contenido de saponinas); en este caso, se recomienda estudiar el grado de discriminación y hacer las comparaciones utilizando otros métodos multivariados.

Otras variables importantes que discriminan suficientemente el germoplasma de quinua son el diámetro de tallo y la altura de planta que ocuparon el quinto y el sexto lugar respectivamente, igualmente el tamaño de panoja que precedió las anteriores.

Análisis de conglomerados. El dendrograma obtenido a través de la distancia euclidiana al cuadrado y el algoritmo de enlace completo muestra que las mejores asociaciones las formaron las variables inicio de floración (IF) y 50% de floración ($F_{(50\%)}$), y diámetro de grano (DG) con el peso de 100 granos (P100), en comparación con el resto de las variables consideradas en el análisis (Figura 1.3).

De acuerdo con el nivel de significancia es posible formar tres grupos de variables claramente definidos, cada uno con varios subgrupos (Recuadro 1.1).

Según los resultados presentados en el dendrograma se puede concluir que si se mejora o modifica la duración del ciclo de la quinua también se modificará su arquitectura de planta.

El comportamiento o la asociación no-significativa entre el primero y el segundo grupo de variables sugiere la posibilidad de combinar en un mismo genotipo los caracteres del grano con los fenológicos y de arquitectura de planta, lo que facilitaría los trabajos de selección sobre accesiones con granos sobresalientes, sin afectar sus características agronómicas.

El comportamiento independiente que tiene la variable emergencia de plántulas se puede atribuir a que ésta depende más de variables externas como la preparación y la humedad en el suelo al momento de la siembra, que de factores genéticos.

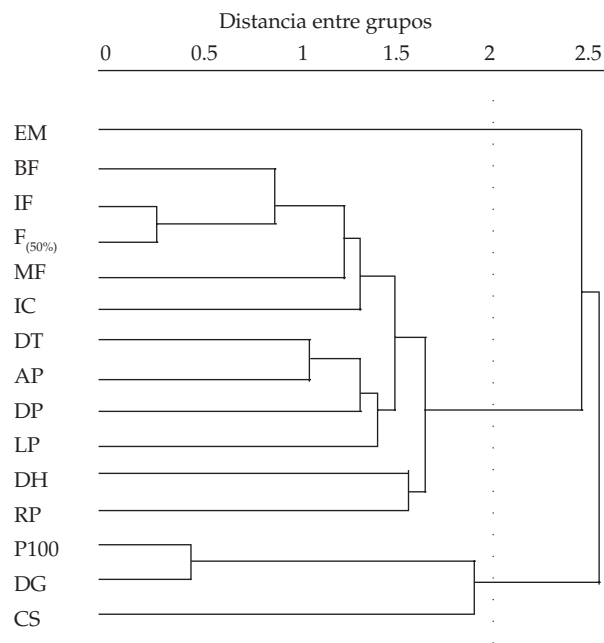


Figura 1.3. Dendrograma de distancias entre 15 variables cuantitativas del germoplasma de quinua. Los nombres de las variables aparecen en la Figura 1.2.

Recuadro 1.1. Grupos y subgrupos resultantes en el agrupamiento de variables en accesiones de quinua.

Grupos
1. Variables fenológicas con las variables de arquitectura de planta
Subgrupos
<ul style="list-style-type: none">• Inicio de floración (IF) con el 50% de floración (F_(50%)) forman la asociación más importante. También se asocian botón floral (BF), la madurez fisiológica (MF) e índice de cosecha (IC).• La asociación más importante la forman el diámetro de tallo (DT) con la altura de planta (AP). Se asocian el diámetro de panoja (DP) y la longitud de panoja (LP) que también se fusionan principalmente con AP la cual, a su vez, se asocia significativamente con IF, F_(50%), BF y MF.• Formado por el número de dientes en las hojas (DH) y el número de ramas principales (RP), que se fusionan significativamente, aunque en menor importancia, con el conglomerado que forman el subgrupo 1 (IF, F_(50%), BF y MF) con el subgrupo 2 (DT, AP, DP y LP).
2. Conformado por las variables de grano. La asociación diámetro de grano (DG) y el peso de 100 granos (P100) es la más importante, a la cual también se asocia, aunque en menor importancia, el contenido de saponina (cs).
3. Formado por la variable emergencia de plántulas (EM), que es la menos importante.

Conclusiones

Los métodos aplicados para estudiar los patrones de variación del germoplasma de quinua fueron complementarios y su contribución fue diferente, como se explica a continuación.

Matriz de correlación . A pesar de que los coeficientes resultantes en este análisis se deben interpretar con precaución, este método ayudó en la identificación del grado de correlación entre las variables, sobresaliendo los coeficientes entre el inicio de floración y el 50% de floración ($r = 0.94$), y el diámetro de grano con el peso de 100 granos ($r = 0.89$) (ver Cuadro 1.1).

Entre las fenológicas se destacaron las correlaciones altamente significativas entre las variables florales y la madurez fisiológica; por otra parte, los coeficientes negativos del índice de cosecha indican que su comportamiento es inversamente proporcional con la floración y la duración del ciclo del cultivo de quinua. La emergencia de plántulas no presentó correlación significativa con otras variables.

Entre las variables de arquitectura de planta se destacaron los coeficientes entre el diámetro de tallo y el diámetro de panoja, entre la altura de planta y el diámetro de tallo, y entre la longitud de panoja y la madurez fisiológica. Este último sugiere que las accesiones de quinua tardías desarrollan plantas altas.

El diámetro de grano y el peso de 100 granos formaron la segunda correlación más importante de la matriz, adhiriéndose significativamente el contenido de saponina con el diámetro de grano. Esta última asociación positiva indica que los granos más grandes presentan un mayor contenido de saponinas.

Análisis de componentes principales. Este método permitió identificar las variables que más contribuyeron a la varianza en cada uno de los tres componentes seleccionados y, para determinar su grado de discriminación o importancia (ver Cuadro 1.2 y Figura 1.1).

El primer componente se caracterizó principalmente por las variables fenológicas y el índice de cosecha, este último caracterizó en sentido contrario; también lo hicieron en forma secundaria la altura de planta y el diámetro de tallo, indicando que el germoplasma dispone de una amplia variabilidad en estas características.

El segundo componente lo constituyeron las variables del grano que caracterizaron en sentido contrario por la longitud de panoja, indicando que en estas características también existe una amplia variabilidad.

El tercer componente estuvo formado por las variables diámetro de tallo, diámetro de panoja y la altura de planta, y en forma secundaria, por las variables del grano. Las demás variables de arquitectura de planta contribuyeron en forma proporcional con estos componentes.

A través de la proporción de la varianza fue posible determinar el grado de discriminación de cada una de las variables. Las variables fenológicas fueron más discriminatorias que las variables del grano y éstas, a su vez, más que las variables de arquitectura de la planta, lo que permite concluir que las variables que tienen relación con la reproducción de la planta son más discriminatorias que las vegetativas.

Análisis de conglomerados. Las asociaciones confirman las variables que fueron identificadas por la matriz de correlación y por el grado de discriminación identificado en el análisis de componentes principales. Considerando la significancia de cada uno de los grupos de variables es posible obtener información de utilidad para el programa de mejoramiento del cultivo de la quinua.

Es importante señalar que no siempre es necesario aplicar los tres métodos para estudiar las variables consideradas, en algunos casos, cuando se conoce bien el germoplasma, es suficiente la aplicación de uno o máximo dos de los métodos descritos para obtener y comprobar resultados válidos.

Referencias

- Crisci, J. V. y López, M. F. 1983. Introducción a la teoría y práctica de la taxonomía numérica. Secretaría General de la Organización de Estados americanos (OEA). Washington, D.C. 132 p.
- Crivisqui, E. 1998. Presentación de los métodos de clasificación. En: Seminario de métodos estadísticos multivariados. Enero 5 a 16, Valdivia, Chile. 56 p.
- Cupolillo, E.; Momen, H.; y Grimaldi, G. Jr. 1998. Genetic diversity in natural populations of new world *Leishmania*. Mem. Inst. Oswaldo Cruz, Rio de Janeiro 93(5):663-668.
- Dillon, W. R. y Goldsten, M. 1984. Multivariate analysis: Methods and applications. Wiley. Nueva York 587 p.
- Ferreira, P. 1987. Análisis multivariado aplicado a problemas de clasificación y tipificación. En: Taller sobre aplicaciones del análisis multivariado. Instituto de Educación Continuada (IDEC). Antigua. 12 p.
- Hair, J. F.; Anderson, R. E.; Tatham, R. L.; y Black, W. C. 1992. Multivariate data analysis. Macmillan Publ. Co., Nueva York. 544 p.
- López J. A. e Hidalgo, M. D. 1994a. Análisis de componentes principales y análisis factorial. En: Ato, M. y López, J. J. (eds.). Fundamentos de estadística con Systat. Addison Wesley Iberoamericana. p. 457-503.
- López J. A. e Hidalgo, M. D. 1994b. Análisis de conglomerados. En: Ato, M. y López, J. J. (eds.). Fundamentos de estadística con Systat. Addison-Wesley Iberoamericana. p. 505-532.
- Pla, L. E. 1986. Análisis multivariado: Método de componentes principales. Secretaría General de la Organización de los Estados Americanos (OEA). Washington, D.C. 95 p.
- Plucknett, D. L.; Williams, J. T.; Smith, N.; y Anishetty, N. N. 1992. Los bancos genéticos y la alimentación animal. Editorial IICA, San José, Costa Rica. 257 p.
-

Caso 2. Análisis de la variabilidad genética en fríjol

*Gustavo Ligarreto**

El fríjol tiene una amplia variabilidad genética y sus caracteres morfológicos, especialmente el color de las semillas y el hábito de crecimiento, han sido utilizados para la selección de variantes. Esta selección y la deriva genética han sido los principales factores evolutivos responsables del mantenimiento y la ampliación de la variación morfológica. Los plantas y los granos de fríjol se identifican por el tamaño, el color de las semillas, el hábito de crecimiento y la adaptación en regiones con condiciones de clima adversas (Ligarreto, 1991; Angulo y Arcila, 1990).

Colombia posee una extensa colección de fríjol resultante del trabajo de mejoramiento bajo condiciones de adaptación a clima frío (2000 a 2700 m.s.n.m.) en la zona Andina, formada por 165 accesiones regionales e introducidas de los hábitos I, II y III con crecimiento arbustivo e indeterminado que han sido progenitores de variedades mejoradas y líneas segregantes y élites. A pesar de que dicha colección ha sido utilizada como fuente de genes para el fitomejoramiento de la especie, en la actualidad se tiene poca información acerca de su constitución genética y aún existen dudas de la variabilidad de los nuevos cultivares mejorados. Por tanto, su caracterización permitirá el uso menos empírico de los materiales en los programas futuros de mejoramiento así como el conocimiento suficiente de las fuentes de los genes que se deben combinar y conservar.

El objetivo en este ejemplo es presentar un caso frecuente entre curadores y usuarios de los bancos de germoplasma, que consiste en conocer la variabilidad genética de fríjol mediante el uso de marcadores morfológicos y agronómicos, utilizando un análisis unificado de marcadores.

De la colección en estudio se seleccionaron como testigos tres variedades mejoradas y comerciales, una línea élite y 26 accesiones tomadas al azar con el objeto de extrapolar los resultados de variabilidad a la población original; esta muestra representó el 18% de la población total siendo superior a la considerada en la mayoría de otros estudios, que sólo alcanza 10% de la población.

Estos materiales fueron sembrados en el Centro de Investigación Tibaitatá (1900 m.s.n.m.) de la Corporación Colombiana de Investigación Agropecuaria (Corpoica) y en el municipio de Calima-Darién (1200 m.s.n.m.), Valle del Cauca, Colombia, en seis ambientes y diferentes lotes y semestres, en un diseño experimental de bloques completos al azar con 30 tratamientos (accesiones) y cuatro replicaciones. La unidad experimental, de 9.6 m², estaba constituida por cuatro surcos de 4 m de longitud cada uno, distanciados 0.6 m entre ellos y 10 cm entre plantas.

En total se evaluaron 23 caracteres cuantitativos correspondientes a los diferentes estados fenológicos de la planta y relacionados con el rendimiento y sus componentes: Cinco en estado de plántula, siete en floración, uno en estado de madurez fisiológica y 10 en época de cosecha; y 42 caracteres cualitativos conformados por 13 binarios y 29 categóricos en los estados fenológicos de plántula, floración, madurez fisiológica y cosecha (Cuadro 2.1).

* Profesor Asociado. Universidad Nacional de Colombia, Facultad de Agronomía, Bogotá.

Cuadro 2.1. Descriptores varietales morfológicos empleados en la evaluación de accesiones de frijol.

Variables cualitativas		Variables cuantitativas
Color del hipocótilo	Reacción a mildeo	Días a emergencia
Color de la nervadura de hojas primarias	Reacción a crisomélidos	Longitud del hipocótilo
Color de las alas	Color de las vainas	Longitud del epicótilo
Color del limbo estandarte	Patrón de distribución del color de vainas	Ancho de las hojas primarias
Patrón de distribución del color del limbo estandarte	Forma del corte transversal de vainas	Longitud de las hojas primarias
Venas en la flor	Distribución de vainas en la planta	Longitud del folíolo central (4° nudo)
Color del cuello del estandarte	Color de las vainas	Ancho del folíolo central (4° nudo)
Patrón de distribución del cuello estandarte	Patrón de distribución del color de las vainas	Longitud del tallo principal
Color del cáliz	Perfil de las vainas	Altura de coberturas
Color de las bractéolas	Tipo de ápice de la vaina	Número de nudos
Forma de las bractéolas	Grado curvatura del ápice de la vaina	Longitud de las vainas
Tamaño de las bractéolas	Dirección de la curvatura del ápice de la vaina	Ancho de las vainas
Hábito de crecimiento	Consistencia de la vaina	Longitud del ápice de vainas
Color del tallo	Color primario de las semillas	Vainas por planta
Pubescencia del tallo	Color secundario de las semillas	Semillas por vaina
Color de las hojas	Distribución del color secundario	Area foliar del folíolo central
Forma del folíolo central	Aspecto de la testa	Rendimiento por planta
Tipo de ramificación	Venas en la semilla	Rendimiento por parcela
Volcamiento	Color alrededor del hilo	Días a floración
Reacción a roya	Forma de las semillas	Días a madurez fisiológica
	Acervo genético	Días a cosecha
		Peso de 100 semillas

Para determinar los centros de domesticación de las colecciones se evaluaron las características de la flor –color, forma y tamaño de las bractéolas– la forma del folíolo central de la hoja, la presencia de venas en el estandarte y el tamaño de las semillas típicas de los acervos centroamericano y andino de Sur América, que fueron propuestas por Singh et al. (1991a).

Los estados de los descriptores cualitativos utilizados para la caracterización y la evaluación de los insectos-plaga y las enfermedades se calificaron de acuerdo con las escalas de los sistemas estándar para la evaluación de germoplasma de frijol propuestas por el CIAT (Schoonhoven y Pastor-Corrales, 1991) y el registro de los colores se hizo con la tabla de colores Munsell para tejidos vegetales (Muñoz et al., 1993).

Según los objetivos del estudio, los datos de caracterización se evaluaron de las formas siguientes: (1) Mediante el análisis de componentes principales de las variables cuantitativas para determinar cuáles de ellas discriminan la población de cultivares. (2) Por medio del análisis de la varianza genética, ambiente, y la interacción genotipo por ambiente para las variables cuantitativas con el fin de determinar el aporte de estas varianzas al fenotipo. De esta manera, se seleccionaron las variables de mayor heredabilidad para discriminar las accesiones y se hizo la selección de los progenitores. (3) Cuando el objetivo era conocer la variabilidad genética de la colección por parámetros cualitativos y cuantitativos, se unificaron los datos originales tal como se tomaron los registros y se analizaron de manera unificada a través de un mismo coeficiente de

similitud. A continuación se presentan los análisis de los datos utilizando estas tres modalidades de caracterización las cuales, a su vez, pueden ser excluyentes.

Análisis de la variabilidad genética por componentes principales

Inicialmente se calcularon los promedios, las varianzas, las desviaciones estándar y los coeficientes de variación de los caracteres cuantitativos en los siete ambientes donde se realizaron los estudios. Los coeficientes de variación fueron superiores a 8%, lo cual indica que existe un amplio rango de dispersión en la colección de frijol en estudio para cada descriptor, en especial para el rendimiento por parcela que presentó el valor más alto (80.4%). Las varianzas y desviaciones estándar para rendimiento por parcela y días a cosecha mostraron los valores más altos con respecto a los de los demás caracteres.

El número de observaciones para establecer el grado de asociación entre las 23 variables cuantitativas evaluadas fue alto, entre 270 y 760, lo cual favoreció la significancia para los bajos coeficientes de correlación encontrados. Debido a la existencia de una alta correlación entre las variables, fue necesario aplicar las estadísticas multivariadas como componentes principales con el fin de conocer cuáles de ellas discriminan las accesiones de frijol en estudio.

Cada componente tiene un vector característico asociado conformado por los coeficientes de todas las variables consideradas en el estudio, por tanto, se seleccionaron los coeficientes con mayor valor absoluto que corresponden con las variables que más aportan en la discriminación varietal (Recuadro 2.1).

Recuadro 2.1. Componentes y coeficientes de las variables importantes en la discriminación varietal de una colección de frijol.

Componentes	Variables	Valor del coeficiente
1	Longitud de epicótilo	0.267
	Ancho de hojas primarias	0.275
	Longitud de hojas primarias	0.268
	Número de nudos	-0.278
	Número de vainas	-0.289
	Días a floración	-0.283
	Días a madurez fisiológica	0.266
	Peso de 100 semillas	0.251
2	Longitud del folíolo central	0.326
	Ancho del folíolo central	0.348
	Area del folíolo central	0.356
3	Días a emergencia	-0.586
4	Longitud del hipocótilo	-0.564
5	Número de plantas cosechadas por parcela	-0.564
6	Número de semillas por vaina	0.447

En el análisis de componentes principales (Cuadro 2.2) se puede observar que los primeros seis poseen un valor característico > 1 , siendo los de mayor importancia en el estudio, con una representación del 87.31% de la variación total. El primer componente contribuyó con 43.86% y el segundo con 19.64%, mientras que el sexto contribuyó con

Cuadro 2.2. Componentes principales seleccionados por su valor característico ($\lambda \geq 1$) a partir de 23 variables morfológicas cuantitativas evaluadas en siete ambientes.

Componentes principales ^a	Valores característicos (λ)	Proporciones de la varianza total	
		Absoluta (%)	Acumulada(%)
1	10.089	43.86	43.86
2	4.518	19.64	63.51
3	1.859	8.08	71.60
4	1.431	6.22	77.82
5	1.143	4.97	82.79
6	1.038	4.51	87.31

a. La definición de los componentes aparece en el Recuadro 2.1.

4.51% de esta variación. El alto porcentaje de la variación total explicada por los primeros seis componentes sugiere que estos contienen variables que discriminan bien la colección de frijol.

Este análisis permitió identificar que las variables relacionadas con hábito de crecimiento, precocidad, área foliar y rendimiento por planta están asociadas con los seis componentes de mayor importancia entre los 23 posibles. De igual manera, se identificaron el número de nudos, días a floración, tamaño de la hoja, número de semillas por vaina y rendimiento por planta como los caracteres que más contribuyen para la discriminación de los cultivares en los acervos genéticos mesoamericano y andino.

Los resultados anteriores concuerdan con los encontrados por Singh et al. (1991b) quienes hallaron una alta asociación entre los acervos genéticos con las características consideradas en este estudio, tales como hojas y semillas grandes en individuos del acervo andino y pequeñas en el acervo mesoamericano. Voysest (1991) sugiere que el tamaño de las semillas es uno de los caracteres de mayor utilidad en la clasificación de frijol, ya que determina el grado de aceptación de las variedades tanto por parte del productor como del consumidor.

La representación gráfica de los componentes principales (Figura 2.1) muestra que en los cultivares andinos y mesoamericanos existen amplias diferencias morfológicas y agronómicas, las cuales son mayores entre los primeros que presentan un mayor grado de dispersión en los tres primeros componentes. La diversidad entre los cultivares andinos está representada por los hábitos de crecimiento determinado e indeterminado con separación entre arbustivos determinados para los materiales de Colombia, México y Bolivia; peruanos de hábito indeterminado; y una accesión (Argentina 2) procedente del sur de los Andes con hábito de crecimiento determinado, semillas de color negro y características morfológicas parecidas a las del subgrupo de andinos determinados. No obstante, esta última se diferencia de este subgrupo por presentar germinación tardía (30 días o más), aspecto que la aproxima a algunos materiales silvestres.

Las variantes dentro del acervo andino se deben a la presión de selección por parte de los productores que prefieren variedades de frijol con características fenotípicas que tienden hacia el predominio de ciertas variedades a nivel local o regional. Sin embargo, cultivares morfológicamente similares pueden ser distantes en su constitución genética.

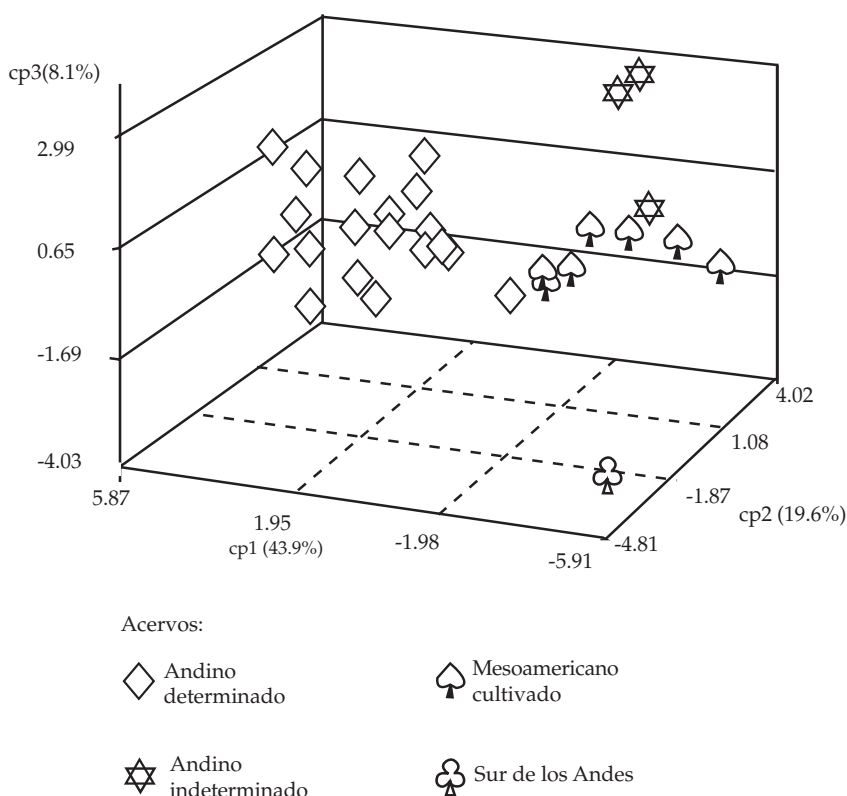


Figura 2.1. Dispersión de 30 accesiones de la colección colombiana de frijol en los tres primeros componentes principales de 23 descriptores cuantitativos. Las accesiones andinas indeterminadas son Perú 5, Tundama y Perú 224; el cultivar Argentina 2 constituye el grupo sur de los Andes.

Variabilidad representada por variables cuantitativas de alto efecto genético

Otra forma de agrupar las accesiones por su morfología consiste en utilizar los descriptores que tienen alta heredabilidad los cuales, a su vez, aseguran alta estabilidad en diferentes ambientes. Para el efecto, en el estudio de este ejemplo se hallaron los estimadores de los componentes de varianza de los 23 descriptores cuantitativos en los siete ambientes de evaluación y con ellos se obtuvieron los coeficientes de repetibilidad o heredabilidad (γ)¹ (Goodman y Paterniani, 1969). El valor más alto de este coeficiente correspondió al peso de 100 semillas ($\gamma = 6.04$) seguido por la longitud de las vainas ($\gamma = 4.83$). El número de nudos ($\gamma = 1.07$) y de vainas por planta ($\gamma = 1.08$), la longitud del ápice de

$$\gamma = \frac{\sigma_g^2}{\sigma_a^2 + \sigma_{ga}^2}$$

1. Heredabilidad: $\gamma = \frac{\sigma_g^2}{\sigma_a^2 + \sigma_{ga}^2}$. Los valores $\gamma > 1$ indican que la medida entre accesiones tiene más expresión genética que efecto del ambiente e interacciones. Si $\gamma < 1$, la heredabilidad es baja y existe efecto del ambiente sobre la expresión del carácter (Goodman y Paterniani, 1969).

las vainas ($\gamma = 1.33$) y la época a madurez fisiológica ($\gamma = 1.01$) presentaron valores cercanos a la unidad. Los valores < 1 significan diferencias entre las accesiones a través de los ambientes para el carácter en medición, tal como ocurrió con la repetibilidad en el rendimiento por parcela ($\gamma = 0.02$), edad a floración ($\gamma = 0.05$), edad a cosecha ($\gamma = 0.05$) y número de plantas por parcela ($\gamma = 0.003$).

La consistencia en la selección de las variables por su alto valor de repetibilidad para discriminar la variabilidad morfoagronómica es confirmada por los altos coeficientes del valor absoluto encontrados en el análisis de componentes principales.

Las seis variables continuas seleccionadas presentaron un alto grado de asociación entre sí, por tanto, fueron analizadas mediante componentes principales, encontrándose en los dos primeros valores característicos > 1 que explican el 85.38% de la variación total -67.31% para el primer componente y 18.07% para el segundo-.

Con los coeficientes de los primeros dos componentes principales se construyó el dendrograma de la Figura 2.2, en el cual en cinco unidades de distancia euclidiana aparecen tres grupos de accesiones con sus respectivos subgrupos de discriminación,

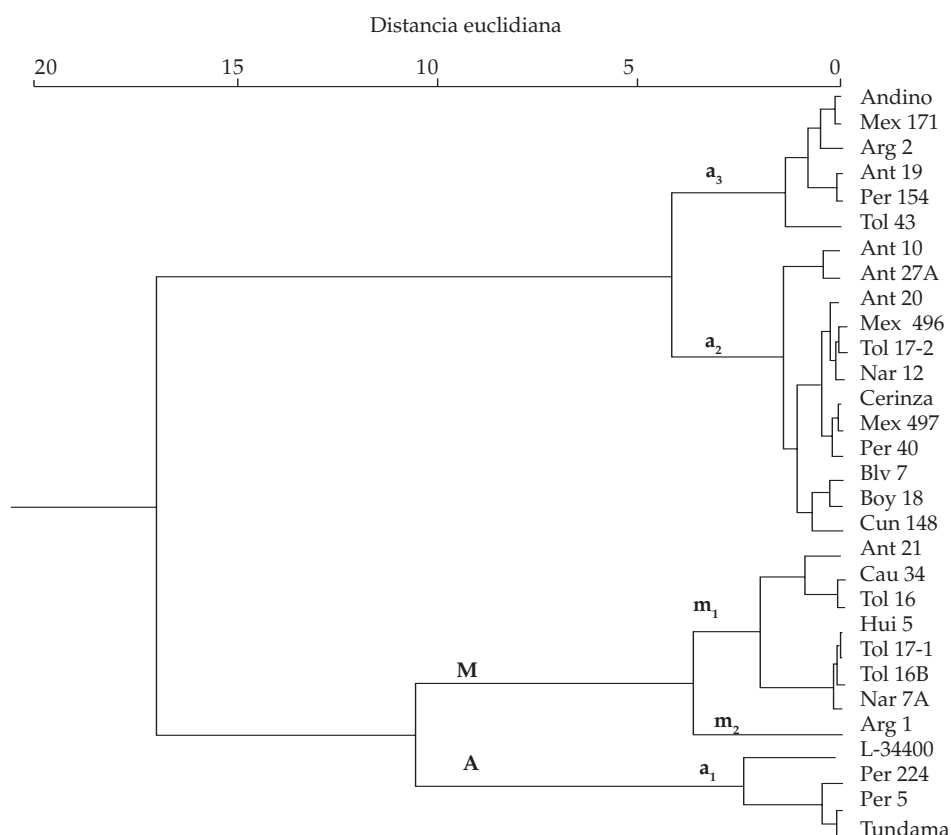


Figura 2.2. Dendrograma de 30 accesiones de frijol construido a partir de seis variables cuantitativas seleccionadas usando el coeficiente de heredabilidad $\gamma > 1$.
Acervos: A = andino, subgrupos a_1 , a_2 , a_3 , M = mesoamericano, subgrupos m_1 , m_2 .

principalmente por acervo genético andino y mesoamericano. El subgrupo andino a_1 se encuentra cerca a los mesoamericanos con hábito de crecimiento indeterminado tipos II y III, excepto la accesión L-34400 que tiene crecimiento determinado, pero con alta producción de vainas por planta (entre paréntesis) (20.1). Este último carácter también discrimina este subgrupo y los mesoamericanos (22.8 y 27.3) de los otros andinos ($a_2 = 12.4$ y $a_3 = 15.0$) (Cuadro 2.3).

El acervo mesoamericano contiene dos subgrupos diferenciados por los tiempos tardíos de germinación y madurez fisiológica: El m_2 constituido por Argentina-1, del sur de los Andes, de hábito de crecimiento tipo II que presentó 22.5 días a emergencia y 142 días a madurez fisiológica; y el subgrupo m_1 , procedente del norte de los Andes, de crecimiento indeterminado tipo III –con excepción de Tolima17-1 de hábito de crecimiento tipo II– que presentó 16.8 días a emergencia y 136 días a madurez fisiológica.

En el mismo Cuadro se incluyen otras variables que permitieron clasificar las accesiones por acervo y subgrupos respectivos. Por ejemplo, el peso de 100 semillas fue inferior a 25 g en los subgrupos m_1 y m_2 , entre 25 y 40 g para el subgrupo a_1 y superior a 40 g en los subgrupos de andinos a_2 y a_3 correspondiendo a semillas pequeñas, medianas y grandes, respectivamente (Muñoz et al., 1993).

Cuadro 2.3. Promedios por grupos genéticos de las principales variables cuantitativas en caracterización morfológica de frijol^a.

Variables	Grupos y subgrupos				
	Andinos (A)			Mesoamericanos (M)	
	a_1	a_2	a_3	m_1	m_2
Días a emergencia	16.1	18.1	19.6	16.8	22.5
Longitud de hipocótilo (cm)	4.5	4.6	5.4	4.1	4.5
Ancho folíolo central (cm)	6.7	6.6	6.7	5.7	6.2
Altura de cobertura (cm)	51.4	36.0	34.1	40.5	30.4
Número de nudos	12.4	6.8	8.8	11.2	10.0
Longitud de vaina (cm)	11.8	11.4	11.9	9.6	8.5
Vainas por planta (no.)	20.1	12.4	15.0	2.8	27.3
Longitud ápice de vaina (cm)	1.3	1.6	1.3	1.0	0.6
Semillas por vaina (no.)	3.7	3.7	3.9	4.2	4.0
Area folíolo central (cm ²)	44.1	42.5	43.2	32.8	37.0
Rendimiento por planta (g)	21.8	14.6	15.7	16.6	14.8
Días a floración (no.)	74.0	64.0	66.0	74.0	70.0
Días a madurez fisiológica	139.0	123.0	129.0	136.0	142.0
Peso de 100 semillas (g)	37.5	44.3	41.8	22.2	18.1

a. Los grupos genéticos se encuentran relacionados en la Figura 2.2.

Caracterización genética por descriptores cualitativos y cuantitativos

Cuando los investigadores desean hacer poco manejo estadístico de los datos de descriptores morfológicos cualitativos y cuantitativos, recurren frecuentemente al uso de los promedios o modas de cada carácter por accesión y realizan un análisis unificado de descriptores mediante la distancia de Gower. Este método permite presentar la dispersión de las accesiones en un plano cartesiano o en un dendrograma.

En la Figura 2.2 antes mencionada se presenta la variabilidad genética de las 30

accesiones de frijol, medida en las 65 variables de descripción varietal de tipos morfológico cualitativo y cuantitativo. La variabilidad está representada, a su vez, por tres grupos: (1) típico andino conformado por 16 accesiones con denominación a_2 , (2) integrado por los acervos andino y mesoamericano con separación en subgrupos a_1 y m_1 , y (3) igual al grupo anterior más cultivares andinos (a_3) y un cultivar mesoamericano (m_2). Se observa que estos resultados son similares a los encontrados en el análisis de escalamiento multidimensional.

Teniendo en cuenta algunos de los caracteres de alta heredabilidad y las principales características para la clasificación de razas de frijol (Singh, 1985; Singh et al., 1991a), el grupo correspondiente al acervo andino se caracteriza por el color amarillo, café-rojizo y verde con pigmentación rosada en los cotiledones, el color blanco y morado de las alas de la flor, el hábito de crecimiento tipos I y II, la forma romboédrica del folíolo central, las bractéolas lanceoladas, el color primario rojo y crema-oscuro de las semillas, y la forma ovoide y arriñonada recta de éstas.

El acervo mesoamericano presentó las características morfológicas siguientes: Color amarillo pálido de los cotiledones, alas moradas, hábito de crecimiento tipos II y III, folíolo central de la hoja ovalado, bractéolas de forma cordada, color primario rojo y crema suave de la semilla con formas ovoidea y alargada-ovoidea. Dentro de cada acervo la variabilidad está representada por diferencias varietales, así, dentro del acervo andino, el subgrupo a_1 representado por Perú-5 y Tundama (que tienen en Perú-5 uno de sus progenitores) presenta el estandarte blanco; el estandarte en el grupo a_2 es lila; y el subgrupo a_3 , formado por las accesiones Argentina-2, Bolivia-7, L-43400 y Perú-224 presenta con mayor frecuencia estandarte morado.

Los estandartes en el acervo mesoamericano son blancos para el subgrupo m_1 , constituido por siete accesiones, y morado para Argentina-1 que conforma el subgrupo a_2 . El hábito de crecimiento en los andinos es arbustivo indeterminado (tipo II) en el subgrupo a_2 y determinado (tipo I) en los subgrupos a_1 y a_3 ; mientras que en los mesoamericanos, en el subgrupo m_1 , con siete accesiones, predomina el hábito indeterminado tipo III y en el subgrupo m_2 el crecimiento es de tipo III.

Según Ligarreto y Martínez (2001) y Singh et al. (1991b) el rendimiento, el número de vainas por planta y el peso de 100 semillas son los caracteres que más contribuyen para la descripción de colecciones de frijol. En este caso, las accesiones del acervo andino presentaron un menor número de vainas por planta (12.4 a 20.1) en comparación con los subgrupos del acervo mesoamericano (22.7 a 27.3). Por el contrario, los cultivares del acervo andino mostraron mayores pesos de semillas que los mesoamericanos (37 - 44 g vs. 18 - 22 g) (Cuadro 2.3).

Singh et al. (1991a) en una clasificación de accesiones de frijol cultivado presentes en dos acervos hereditarios y en la que consideraron las características foliares, de flor y de tamaño de las semillas, encontraron una mayor presencia de los caracteres morfológicos representativos del conjunto de gametos del acervo andino, en relación con el acervo mesoamericano. Lo anterior se puede explicar por una mayor representación en la colección de semillas con tamaño entre mediano y grande, y un mayor número de recolecciones en zonas de climas medio y frío colombiano. Por otra parte, como resultado de las regeneraciones sucesivas de la colección a partir de 1929, cuando se inició el programa de mejoramiento de la especie en Colombia, es posible que se hayan

seleccionado accesiones por su tamaño de semillas.

La aplicación del algoritmo UPGMA² a distancias Gower en los datos combinados de 42 variables categóricas y 23 variables continuas, dio como resultados conglomerados similares a los de componentes principales para caracteres cuantitativos (ver Figura 2.1), con separación morfológica y agronómica en los dos acervos ancestrales de domesticación

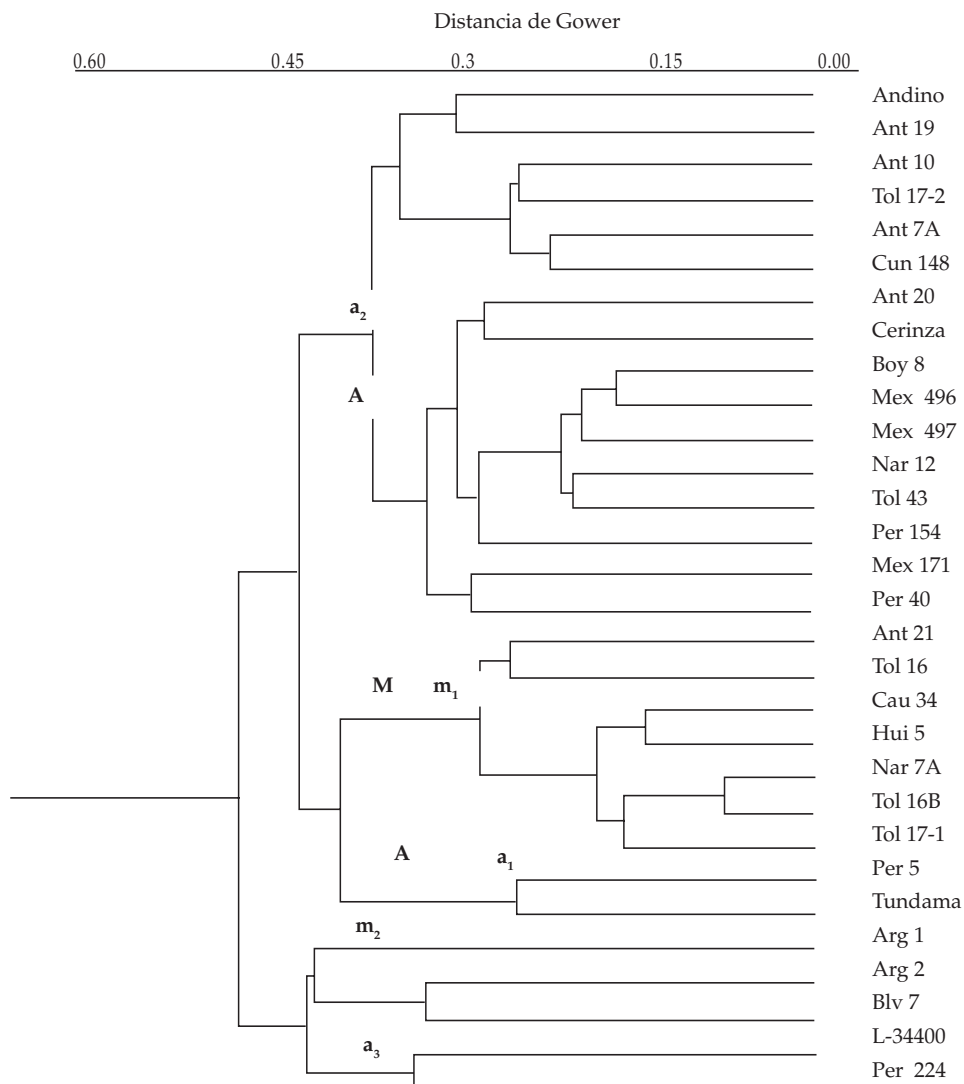


Figura 2.3. Relación morfológica entre 30 accesiones de frijol basada en las modas de 42 descriptores cualitativos y en el promedio de 23 descriptores cuantitativos.
Acervos: A = andino, subgrupos a₁, a₂, a₃; M = mesoamericano, subgrupos m₁, m₂.

2. UPGMA (Unweighted Group Meted Arithmetic Average) compara distancias promedio entre accesiones.

andino y mesoamericano. El acercamiento detectado entre los dos cultivares Argentina, pertenecientes a diferente acervo debido a sus diferencias morfológicas (Figura 2.3), puede ser debido a la expresión de ciertos genes por adaptación ecológica.

El análisis combinado de los datos de caracteres cualitativos y cuantitativos, con agrupamientos por métodos multivariados, proporciona un panorama amplio del conocimiento de la variabilidad genética que permite su utilización en programas de mejoramiento varietal y en la identificación de descriptores confiables de las accesiones.

La principal ventaja de esta propuesta es la flexibilidad que ofrece para la aplicación de medidas apropiadas a cada tipo de datos en la colección de germoplasma, como también para la inclusión de información sobre el origen geográfico de las especies (Cole-Rodgers et al., 1997). Los datos naturales utilizados en el análisis son incorporados en una sola matriz conjunta con el objeto de realizar un análisis completo y establecer las posibles relaciones entre todos los cultivares mediante los descriptores evaluados simultáneamente.

Referencias

- Angulo N. F. y Arcila, M. B. 1990. Factores que limitan la productividad del frijol voluble en Nariño. Informe técnico no. 59. Instituto Colombiano Agropecuario (ICA), Regional 5. Pasto, Nariño, Colombia. 23 p.
- Cole-Rodgers, P.; Smith, D. W.; y Bosland, P. W. 1997. A novel statistical approach to analyze genetic resource evaluations using *Capsicum* as an example. *Crop Sci.* 37:1000-1002.
- Goodman, M. M. y Paterniani, E. 1969. The races of maize III. Choices appropriate characters for racial classification. *Econ. Bot.* 23:265-273.
- Ligarreto, G. A. 1991. Consideraciones generales sobre el cultivo de frijol en Colombia. *Revista ICA* 26:1. 25 p.
- Ligarreto, G. A. y Martínez, O. 2001. Variabilidad genética en frijol común (*Phaseolus vulgaris* L.). 2. Análisis de marcadores morfológicos y agronómicos cualitativos y cuantitativos. (s.p.).
- Muñoz, G.; Giraldo, G.; y Fernández de Soto, J. 1993. Descriptores varietales: Arroz, frijol, maíz, sorgo. Publicación no. 177. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia. 170 p.
- Singh, S. P.; Gepts, P.; y Debouck, D. G. 1991a. Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot.* 45(3):379-396.
- Singh, S. P.; Gutiérrez, A., Molina, A., Urea, C.; y Gepts, P. 1991b. Genetic diversity in cultivated common bean. 2. Marker-based analysis of morphological and agronomic traits. *Crop Sci.* 31:23-29.
- Singh, S. P. 1985. Conceptos básicos para el mejoramiento del frijol por hibridación. En: M. López; F. Fernández y A. Van Schoonhoven (eds.). *Frijol investigación y producción*. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia. p. 109-126.
- Schoonhoven, A. y Pastor-Corrales, M. A. 1991. Sistema estándar para la evaluación de germoplasma de frijol. Centro Internacional de Agricultura Tropical (CIAT), ISBN 84-89206-73-2. 56 p.
- Voysest, O. 1991. Bean Cultivars: Classes and commercial seed types. En: Van Schoonhoven, A. y Voysest, O. (eds.). *Common Beans: Research for crop improvement*. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia. p. 124.
-

Caso 3. Análisis de la Variabilidad genética de jícama

César Tapia*

En el estudio de este caso se utiliza como ejemplo la caracterización de la colección de jícama (*Pachyrhizus tuberosus* (Lam.) Spreng) formada por 31 accesiones cultivadas del banco de germoplasma en el CATIE, que considera la variabilidad a nivel del centro de origen (Suramérica). Los genotipos fueron sembrados en un diseño completamente al azar con 35 repeticiones y se caracterizaron morfológicamente utilizando 70 descriptores cualitativos y cuantitativos.

Resultados

Tamaño de muestra. En la investigación se utilizó una probabilidad del 5%, es decir, cuando en el 95% de los casos el promedio de la muestra se encontraba dentro del rango del 5% de la media verdadera de la población. Se considera que esta probabilidad es aceptable en este tipo de estudios. Para mantener un nivel de confianza adecuado se seleccionaron límites de error o confianza $< 25\%$.

Para los caracteres cualitativos se seleccionaron tamaños mínimos entre una y tres muestras. En el caso de los caracteres cuantitativos la variación fue alta ya que existen descriptores influenciados fuertemente por el ambiente. En el Cuadro 3.1 es posible observar los tamaños de muestra de algunas variables de acuerdo con su variación en la colección en estudio. En la práctica es dispendioso tomar, por ejemplo, 1363 o 341 muestras del grosor de la corteza de la raíz, por tanto, se recomienda tomar el mayor número posible de ellas con la rigurosidad estadística que se elija previamente (5% o 10%).

Cuadro 3.1. Estimación del tamaño de muestra mínimo para variables cuantitativas en la colección de jícama (*Pachyrhizus tuberosus* (Lam.) Spreng.).

Variables cuantitativas	Límite de error	
	5%	10%
	Tamaño de muestra ($P < 0.05$)	
Grosor de la corteza de la raíz comestible (mm)	1363	341
Número de vainas por planta	781	195
Longitud del folíolo principal de la hoja (cm)	594	148
Días a la floración	85	21
Días a la madurez fisiológica de la vaina	74	18
Ancho de la vaina (cm)	50	12

* Instituto Nacional de Investigación Agropecuaria (INIAP), Departamento Nacional de Recursos Fitogenéticos y Biotecnología (DENAREF), Quito, Ecuador

Estimación de la variabilidad genética. El promedio, la desviación estándar (DE), el coeficiente de variación (CV) y el índice de herencia (IH) calculados en algunos descriptores cuantitativos (Cuadro 3.2) permiten una estimación preliminar de la variabilidad de la colección. Los resultados indican que existe alta variación como producto de la influencia del ambiente, por ejemplo, en el grosor de la corteza de la raíz y en el número de vainas; por el contrario, descriptores como ancho de la vaina y días a la madurez fisiológica presentan CV relativamente bajos. Además, los altos valores del IH en los descriptores antes mencionados sugieren que la expresión fenotípica es una buena indicación de su patrón genético.

Cuadro 3.2. **Parámetros para la estimación de la variabilidad genética en la colección de jícama (*Pachyrhizus tuberosus*) (Lam.) Spreng.**

Variables cuantitativas	Parámetros			
	Promedio	DE	CV	IH
Grosor de la corteza de la raíz (mm)	0.9	0.6	66.7	0.05
Número de vainas por planta	17.8	13.2	74.2	0.21
Longitud del folíolo principal de la hoja (cm)	11.3	2.5	22.1	0.26
Días a la floración	103.3	42.9	41.5	0.80
Días a la madurez fisiológica de la vaina	207.5	36.8	17.7	0.81
Ancho de la vaina (cm)	1.9	0.4	21.1	0.78

DE = Desviación estándar. CV = Coeficiente de variación. IH = Índice de herencia.

Con esta información es posible analizar parcialmente la variabilidad genética de la colección, pero aún se necesita otro tipo de análisis que permita medir con mayor precisión dicha variación e identificar posibles duplicados, que en el caso de colecciones grandes contribuyen en forma significativa para la conformación de colecciones núcleo. Como resultado de la cantidad de variables que se utilizan en la caracterización, el análisis multivariado es adecuado y valioso para la identificación de duplicados y relaciones de similitud entre las accesiones, así como para discriminar cuáles son las variables y qué información adicional proporcionan para una futura evaluación.

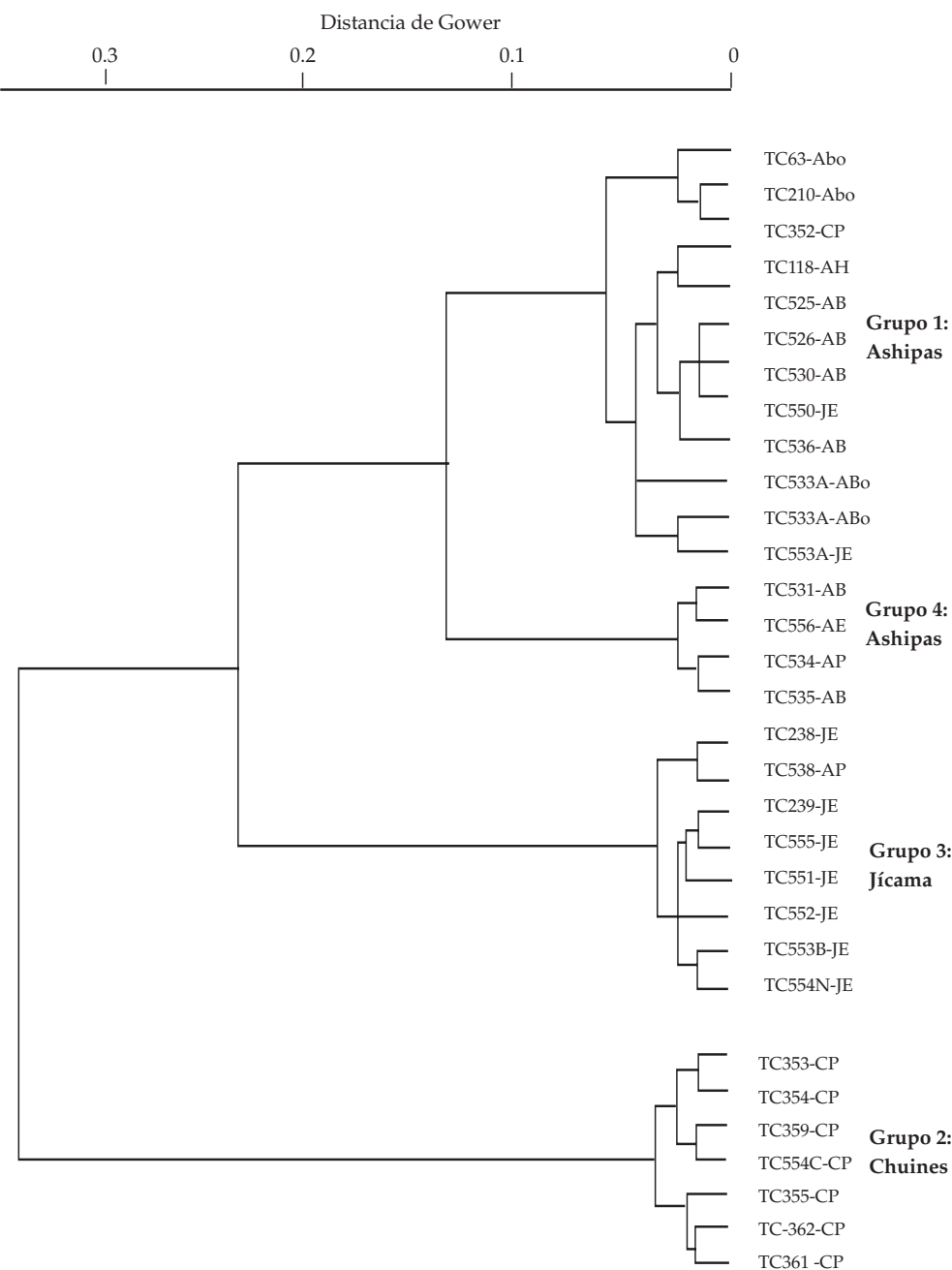
Medición de la variabilidad genética de las colecciones e identificación de duplicados

Mediante el análisis multivariado utilizando la distancia de Gower y el agrupamiento jerárquico de Ward es posible caracterizar adecuadamente colecciones de germoplasma de interés. Los resultados que se obtienen en estos análisis pueden ser de los tipos siguientes (Figura 3.1):

- Agrupamiento de accesiones similares y sus relaciones. En el ejemplo de este caso, se observa que el agrupamiento de Ward produjo cuatro grupos diferentes: las accesiones en el grupo 1 corresponden a las denominas ashipas, el grupo 2 a chuines, el grupo 3 a jícamas y el grupo 4 a accesiones ashipas (TC 531, 556 AE, 534 AP y 535 AB).

- Identificación de duplicados. Es posible asumir que dos accesiones son idénticas morfológicamente si la distancia entre ellas es igual a 0. En el presente caso no se observan duplicados ya que todas las accesiones están en mayor o menor grado distanciadas unas de otras. De todas maneras, se puede observar que existen entradas que morfológicamente son muy similares, como por ejemplo, TC361 con TC362 cuya distancia es muy pequeña (0.01).
- Detección de mutaciones de etiqueta. Se refiere a ciertos errores que se producen en la identificación del número de accesión. En este estudio se observa que la accesión TC352, identificada como chuino de Perú, realmente es ashipano, la accesión TC550, que es ashipano, fue identificada erróneamente como jícama de Ecuador.
- Agrupamientos que permiten diferenciar las accesiones por área geográfica. En el dendrograma se observa que las chuinos son endémicos de la Amazonia del Perú, así como las jícamas se encuentran exclusivamente en la provincia de Manabí en el Ecuador.
- Supuestos sobre la distribución natural de la especie o especies en estudio. En este caso de acuerdo con el dendrograma se puede asumir que las chuinos y las jícamas, que presentan pequeñas variaciones en su morfología, son el resultado de una única introducción desde comunidades vecinas. Con las ashipanos, en cambio, es más complicado hacer supuestos, ya que la dispersión de estas accesiones desde su centro de origen pudo ocurrir por vía fluvial como parece que sucedió en Brasil pero, también, por la recolección e intercambio de los materiales entre bancos de germoplasma y las comunidades indígenas.
- Establecimiento de relaciones y diferencias morfológicas entre grupos. Una vez realizada la identificación de los descriptores discriminatorios para diferenciar grupos de accesiones, mediante el análisis de frecuencias se pueden detectar las características más relevantes que diferencian a cierto grupo de entradas. Los resultados obtenidos indican que las jícamas y las ashipanos presentan la pulpa de la raíz de color blanco, excepto las ashipanos del grupo 4, que al igual que las chuinos, presentan colores amarillos y blanco con morado. La forma de las semillas se presentó como un descriptor interesante para diferenciar entre las chuinos (planas), jícamas (redondeadas) y algunas ashipanos (semiplanas).

Como se mencionó anteriormente, el análisis discriminatorio canónico permite representar espacialmente las accesiones y explicar mediante las variables canónicas (CAN) el porcentaje de variabilidad en cada una de ellas. Además, proporciona información sobre la distancia entre los grupos conformados por el agrupamiento de Ward. En la Figura 3.2 se observa la representación espacial de los grupos. Del análisis se desprende que la variable CAN-1 explica el 99% de la variabilidad total y separa las chuinos y jícamas de los demás grupos; mientras que la variable CAN-2, que explica el 1% de la variabilidad, separa los grupos ashipanos de los otros dos. Se observa que las mayores distancias ocurren entre los grupos de ashipanos y chuinos (4250 y 3877, respectivamente). Por el contrario, los grupos más relacionados evidentemente fueron los dos ashipanos con una distancia de 394.



A = ashipa; C = chuín; J = jícama;
Bo = Bolivia; P = Perú; H = Haití; B = Brasil; E = Ecuador.

Figura 3.1. Dendrograma obtenido por el agrupamiento jerárquico de Ward de la colección de *Pachyrhizus tuberosus* basado en distancias de Gower.

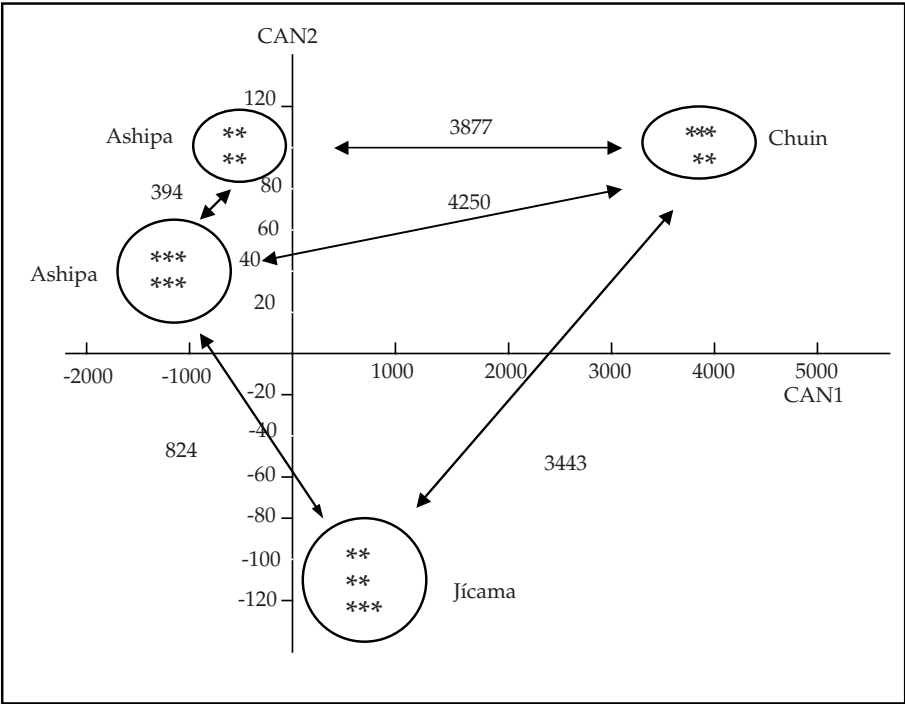


Figura 3.2. Distribución de los individuos en función de las variables canónicas CAN1 y CAN2, procedentes del análisis de resultados del coeficiente de Gower.

Identificación de caracteres cualitativos y cuantitativos discriminatorios entre grupos de accesiones

Caracteres cualitativos

La prueba de ji-cuadrada (χ^2), la prueba de Cramer y el coeficiente de asociación permiten detectar las variables cualitativas que más contribuyen para el establecimiento de diferencias entre grupos genéticos. En el Cuadro 3.3 se observan varios descriptores que discriminan para separar los cuatro grupos, sobresaliendo entre ellos la forma y el tipo del lóbulo central de la hoja con los más altos valores de χ^2 y de asociación. La forma de la semilla también tiene una alta contribución de acuerdo con la prueba de Cramer.

Cuadro 3.3. Caracteres cualitativos de mayor valor discriminatorio entre grupos de accesiones de la colección de *Pachyrhizus tuberosus*.

Carácter	Ji-cuadrada (χ^2)	Coeficiente de asociación (P)	Cramer (V)
Forma del lóbulo del folíolo central de la hoja	59.2**	1.38	0.80
Tipo de lóbulo en el folíolo central de la hoja	53.1**	1.31	0.76
Color de la pulpa de la raíz	47.4**	1.24	0.71
Forma de la semilla	43.3**	1.18	0.89
Celeridad de crecimiento del tallo principal	41.0**	1.15	0.81

** P < 0.01.

Caracteres cuantitativos

Según Engels (1983) un carácter para el cual los grupos presenten valores marcadamente distintos, tendrá un valor discriminante o índice $D_{\text{máx.}} = 1$, ya que las comparaciones posibles serán todas significativas. En el Cuadro 3.4 se observan los cinco caracteres con mayor valor discriminatorio. Tanto los descriptores de hoja y vaina como días a la floración y madurez fisiológica permitieron diferenciar los cuatro grupos, tal como se comprueba con el valor promedio.

Cuadro 3.4. Promedio \pm DE de los valores de caracteres cuantitativos más discriminatorio entre grupos de accesiones de la colección de *Phachyrhizus tuberosus*.

Grupos	Carácter				
	Días a floración	Ancho de folíolo	L/A folíolo	Días maduración	Ancho vaina (cm)
Jíquimas	73.5 \pm 6.00	15.4 \pm 1.41	0.77 \pm 0.01	174.8 \pm 15.99	1.66 \pm 0.09
Ashipas	85.7 \pm 6.31	11.6 \pm 1.33	0.87 \pm 0.05	206.9 \pm 9.88	1.90 \pm 0.11
Ashipas	115.1 \pm 6.85	13.3 \pm 0.62	0.98 \pm 0.02	239.2 \pm 7.79	2.61 \pm 0.05
Chuines	180.1 \pm 4.20	6.6 \pm 0.43	1.81 \pm 0.04	287.0 \pm 5.77	2.06 \pm 0.02
Valor 'D'	1	1	1	1	1

Los descriptores cualitativos y cuantitativos de mayor valor discriminatorio son valiosos en posteriores caracterizaciones con el fin de potenciar el germoplasma de las colecciones, mediante la utilización de variables con una verdadera contribución para la identificación de líneas promisorias de uso inmediato en programas de fitomejoramiento.

Referencia

Engels, J. M. 1983. A systematic description of cacao clones. 1. The discriminative value of quantitative characteristics. *Euphytica* 32:377-385.

Caso 4. Análisis de divergencias interespecíficas con pasifloras andinas

Sergio Segura*

Introducción

Las especies del género *Passiflora* (Passifloraceae) son lianas herbáceas, arbustos o árboles pequeños. Las del subgénero *Tacsonia* son todas plantas perennes, lianas con zarcillos que crecen principalmente sobre arbustos o árboles bajos en matorrales del subpáramo; algunas trepan en el dosel más alto del bosque de altitud media (Escobar, 1988). Este subgénero está geográficamente distribuido en las montañas andinas de Venezuela y al norte de Argentina, siendo un grupo relativamente bien estudiado dentro del género *Passiflora*. La clasificación actual de este género fue realizada por Escobar (1988) y Holm-Nielsen et al. (1988) y en ella se reconocen 10 secciones basadas en las diferencias de las características de estipulas, hojas, brácteas y flor. *Passiflora antioquiensis* Karst y *P. bracteosa* Planch. y Lind. ex Tr. y Planch., que fueron consideradas fuera del subgénero *Tacsonia* en las clasificaciones anteriores, se encuentran actualmente incluidas en este subgénero. *Passiflora manicata* (Juss.) Pers. (subgénero *Manicata*) aparece relacionada con el subgénero *Tacsonia*. Según la mayor parte de los taxónomos, incluyendo a Killip (1938), *Passiflora x rosea* (Karst.) Killip es una primera generación híbrido entre *P. tripartita* var. *mollissima* (Kunth) Holm-Nielsen y Jørgensen y *P. pinnatistipula* Cav. Entre las especies frutícolas, *P. tripartita* var. *mollissima*, o curuba de Castilla, es la más conocida. Esta especie ha sido estrechamente relacionada con otras de la sección *Bracteogama*, como por ejemplo *P. cumbalensis* (Karst.) Harms, aunque tiene un fuerte parecido morfológico con *P. mixta* L. de la sección *Tacsonia*. Escobar (1988) explica esta fuerte similitud entre *P. tripartita* var. *mollissima* y *P. mixta* por un origen evolutivo común, pero las clasifica en secciones diferentes, no obstante, se supone que su clasificación tiene una base filogenética. *Passiflora tarminiana* Coppens y Barney (n.p.), conocida en Colombia como Curuba India, ha sido descrito recientemente como una nueva especie (Coppens et al., n.p.).

Debido al interés existente por estas especies frutícolas y la necesidad de proporcionar materiales mejorados a los productores se han realizado cruces entre *P. tripartita* var. *mollissima* con otras especies de *Passiflora*, pero sin previo conocimiento sobre la variabilidad intraespecífica y las afinidades entre especies (Schoëninger, 1986; Escobar, 1985). Recientemente, instituciones de Venezuela, Colombia, Ecuador, Perú y Bolivia iniciaron un proyecto en colaboración con el IPGRI financiado con fondos del Banco Interamericano de Desarrollo para recolectar, caracterizar y conservar las especies de pasifloras andinas. Como parte de este proyecto regional se diseñó por primera vez un conjunto de descriptores morfológicos para pasiflora que fueron compilados por el Programa Nacional Colombiano de Recursos Genéticos como un primer paso en la caracterización de especies de *Passiflora* (Lobo, 1997). Villacis et al. (1998) caracterizaron 40 individuos de cinco especies del subgénero *Tacsonia* y cinco accesiones de *P. manicata* (subgénero *Manicata*) usando estos descriptores. El análisis de los caracteres cualitativos produjo resultados consistentes con la taxonomía publicada, pero no reveló la variación

* Diversidad Genética de Frutales. Universidad Autónoma, Chapingo, Mexico

intraespecífica. Los descriptores cuantitativos mostraron más variación, pero aquella interespecífica no fue consistente con la taxonomía debido al peso excesivo de los descriptores vegetativos. Posteriormente, la lista de descriptores fue revisada para incluir caracteres florales.

En este caso se presenta la primera parte de un estudio de variabilidad en especies del subgénero *Tacsonia*. El objetivo inicial fue probar la utilidad del conjunto de descriptores que se presumían apropiados para estudiar las divergencias morfológicas entre las secciones, especies y accesiones en este subgénero. En forma paralela se explican los métodos de análisis utilizados y las técnicas estadísticas asociadas. El análisis tuvo como objetivo principal la revisión de las divergencias entre especies cultivadas y las silvestres relacionadas.

La metodología utilizada con las pasifloras andinas es un ejemplo donde las características morfológicas fueron depuradas antes de utilizar métodos factoriales y de clasificación automática para sintetizar las relaciones entre variables y entre accesiones. Con otros grupos de plantas se puede seguir este camino de manera total o parcial, dependiendo del objetivo que se persiga.

Materiales y métodos

Material vegetal

El estudio morfológico se hizo con 36 accesiones y especímenes del herbario de especies del género *Passiflora*, enfocado en particular en 28 accesiones de 17 especies del subgénero *Tacsonia*. La selección de especies y accesiones de cada una de las 10 secciones del subgénero *Tacsonia* se hizo teniendo en cuenta su distribución geográfica incluyendo, como mínimo, un representante por especie y por sección. Las especies endémicas (por ejemplo, *P. bracteosa*) estaban representadas por una accesión y las especies con una distribución más extendida (por ejemplo, *P. tripartita* var. *mollissima*) por una accesión en cada región. La información sobre las accesiones aparece en el Cuadro 4.1. Para las especies de los subgéneros *Tacsonia* y *Manicata* se consideró el tratamiento taxonómico de Escobar (1988), pero *P. mollissima* fue considerada como una variedad botánica de *P. tripartita* (Holm-Nielsen et al., 1988). Las especies de los otros subgéneros fueron clasificadas según Killip (1938).

Observaciones

En las plantas se observaron 24 caracteres cuantitativos y 23 cualitativos que fueron completados con especímenes de herbario (Cuadro 4.2). Los descriptores se basaron en los utilizados por Villacis et al. (1998), pero además se agregaron los siguientes: permanencia de las brácteas, largo y ancho del pecíolo, número de nectarios sobre el pecíolo, número de inflorescencias por nudo, cubrimiento del hipantio por la bráctea, longitud de flor, diámetro del hipantio, longitud de la corola y la posición de estigma. Estos caracteres fueron medidos en tres especímenes por accesión para calcular las medias o modas, según el caso, y hacer los análisis estadísticos subsecuentes.

Cuadro 4.1. Accesiones de *Passiflora* consideradas en el análisis morfológico de variabilidad en este caso.

Sugénero /Sección	Especie	Origen	Código	Características del sitio de recolección			Recolector
				Longitud	Latitud	m.s.n.m.	
Tacsonia							
Poggendorfia	P. pinnatistipula	Boyacá, Col.	PC01	-73.2067	5.6866	2600	Segura y López, 1997
		Tungurahua, Ec.	PE02	-78.3630	-1.2199	2870	Vega y Villacís, 1997
	Passiflora x rosea	Boyacá, Col	OC01	-75.3838	5.6196	3095	Segura y López, 1997
Colombiana	P. antioquiensis	Antioquia, Col.	QC01	-75.4310	6.6080	2100	Medina y Coppens, 1997
	P. lanata	Cundinamarca, Col.	WC01	-74.3088	4.4423	2750	Segura y López, 1997
	P. crispolanata	Boyacá, Col.	CC01	-73.3842	5.6216	3100	Segura y López, 1997
	P. adulteriana	Santander, Col.	KC01	-73.0027	7.0922	2800	Segura y López, 1997
	P. tenerifensis	Valle, Col.	NC01	-76.0650	3.7316	2900	Segura y Coppens,
Parritana	P. parritae	Caldas, Col.	RC01	-75.1383	5.4250	2800	Escobar, 1980
Fimbriatistipula	P. fimbriatistipula	Cauca, Col.	FC01	-76.2511	2.5378	3300	Escobar, 1980
Anomala	P. bracteosa	Santander, Col.	BC01	-73.0050	7.1219	2450	Segura y López, 1997
Bracteogama	P. tripartita var. mollissima	Táchira, Ven.	MV01	-71.9147	8.1686	2500	Barney, 1998
		Boyacá, Col.	MC02	-73.8183	5.6425	2590	Segura y López, 1997
		Valle, Col.	MC03	-76.1521	3.8456	2614	Segura y Coppens, 1997
		Tungurahua, Ec.	ME04	-78.3630	-1.2199	2870	Vega y Villacís, 1997
		Huancavelica, Perú	MP05	-74.6817	-12.4034	2920	López et al., 1997
	P. taminiana	Boyacá, Col.	IC01	-73.8183	5.6425	2590	Segura y López, 1997
		Tungurahua, Ec.	IE02	-78.3630	-1.2202	2500	Vega y Villacís, 1997
	P. cumbalensis	Boyacá, Col.	UC01	-73.2853	5.6953	2780	Segura y López, 1997
		Cundinamarca, Col.	UC02	-74.3088	4.4423	2630	Segura y López, 1997
		Tungurahua, Ec.	UE03	-78.3630	-1.2199	2870	Vega y Villacís, 1997
Ampullacea	P. ampullacea	Chimborazo, Ec.	LLE01	-79.0116	-1.9205	2750	Vega y Villacís, 1997
Trifoliata	P. trifoliata	Junín, Perú	TP01	-77.9216	-9.3695a	3570	Stein, 1980
Boliviana	P. gracilens	Cuzco, Perú	GP01	-71.5033	-13.2500	2920	Escobar, 1979
Tacsonia	P. mixta	Boyacá, Col.	XC01	-72.6994	6.9142	3050	Segura y López, 1997
		Tolima, Col.	XC02	-75.5386	4.4685	2950	Segura y López, 1997
		Valle, col.	XC03	-76.1521	3.8456	2614	Segura y Coppens, 19--
		Tungurahua, Ec.	XE04	-78.3630	-1.2199	2870	Vega y Villacís, 1997
Manicata							
Manicata	P. manicata	Santander, Col.	AC01	-72.6627	7.1586	2130	Segura y López, 1997
		Boyacá, Col.	AC02	-73.5345	5.6031	2150	Segura y López, 1997
		Valle, Col.	AC03	-76.0751	3.7167	2050	Segura y Coppens, 1997
	P. cf. manicata	Valle, Col.	ZC01	-75.7333	4.0832	2470	Segura y Coppens, 1997
Passiflora							
Incarnatae		Valle, Col.	EC01	-76.1333	4.3833	—	Segura, 1997
Tiliaefoliae		Quindío, Col.	LC01	-75.6677	4.5333	—	Barney y lópez, 1996
Decaloba							
Decaloba		Valle, Col.	SC01	-76.1501	3.8506	2530	Segura, 1997
Psilanthus		Valle, Col.	YC01	-76.1563	3.8419	2720	Segura y Coppens, 1997

La taxonomía para el subgénero *Tacsonia* es la propuesta por Escobar (1988). Para los subgéneros *Passiflora*, *Decaloba* y *Manicata*, según Killip (1938) y MacDougal (1994). Para *P. tripartita* y *P. mollissima* de acuerdo con Holm-Nielsen et al. (1988). Para *P. tarmianiana* de acuerdo con Coppens y Barley (2001).

Cuadro 4.2. **Descriptores morfológicos utilizados en el estudio de la variabilidad de las especies del subgénero *Tacsonia*. La medición de los caracteres se hizo de acuerdo con Villacis et al. (1998).**

Organo	Caracteres cualitativos	Código	Caracteres cuantitativos	Código
Tallo	Forma externa	SES	Distancia entre nudos	SIL
	Presencia de antocianina	SPA	Diámetro	SDL
	Pubescencia	SPU		
Zarcillo	Forma del espiral	TSS	Dist. desde la base al espiral	TLS
	Presencia de antocianina	TAP	Longitud del espiral	TSL
			Diámetro del espiral	TSD
Estípulas	Presencia y permanencia	SPP	Longitud	SLE
	Forma	SSH	Ancho	SWI
	Presencia de antocianina	SPA		
Pecíolo			Longitud	PLE
			Diámetro	PDI
Hoja	Tipo	LTY	No. de glándulas del margen	LNG
	Margen foliar	LFM	Longitud del lóbulo central	LLL
	Forma del ápice	LAS	Ancho	LWI
	Relación ancho-largo	LRW	Distancia invaginación e inserción del pecíolo	LLP
	Presencia de pubescencia	LPP	Angulo formado por dos lóbulos laterales	LAL
	Posición de los nectarios	LNP	Número de nectarios en el pecíolo	LNN
Pedúnculo			No. de inflorescencias por nudo	DIN
			Longitud	DEL
			Diámetro	DDI
Brácteas	Tipo	BTY	Longitud	BLE
	Permanencia	BPE	Ancho	BWI
	Forma	BSH	Cobertura del hipanto	BCH
	Color	BCO		
Flor	Orientación	FOR	Longitud	FLE
	Tipo de corona	FCT	Longitud del hipantio	FHL
	Forma de corola	FCS	Diámetro del hipantio	FHD
	Color de pétalos	FPC	Longitud de la corola	FCL
	Color de sépalos	FSF	Posición de los estigmas	FSP
	Posición de la pubescencia	FPP		
	Forma de los sépalos	FSS		
	Forma de los pétalos	SPS		

Análisis estadístico

Inicialmente se hizo una selección de caracteres y a continuación, se realizó una descomposición de la variación en ambos conjuntos de caracteres para comparar aquella inter e intraespecífica y la del error. Los caracteres con una alta variación del error no fueron incluidos en el paso siguiente de análisis factorial.

El Análisis en Componentes Principales (ACP) se realizó con los descriptores cuantitativos seleccionados. Como este método calcula nuevas variables sintéticas o vectores que se combinan como variables de partida y maximizan la variación entre individuos, sólo se incluyeron los materiales de especies de los subgéneros *Tacsonia* y

Manicata como individuos activos y los subgéneros *Passiflora*, *Decaloba* y *Psilanthus* aparecen como suplementarios. Este artificio permite representar en los planos factoriales los individuos que por sus grandes diferencias distorsionarían la imagen global de accesiones de interés, apareciendo sólo proyectadas sin entrar en los cálculos de los componentes.

Se generaron correlaciones de los producto-momentos entre las accesiones para las variables seleccionadas con los datos regularizados. Igualmente se calcularon las partes de la variación total explicada por las nuevas variables (los componentes principales) y se trazaron las proyecciones de las accesiones en los primeros dos componentes del ACP. Para el conjunto de descriptores cualitativos seleccionados se utilizó un tratamiento similar con un Análisis Factorial de Correspondencias (AFC). Las 31 accesiones de las secciones de los subgéneros *Tacsonia* y *Manicata* fueron analizadas como individuos activos y las accesiones restantes aparecen como los individuos suplementarios. En el caso del AFC se aplicaron las mismas reglas de interpretación que para el ACP.

Se estimó la correlación entre los descriptores cuantitativos y cualitativos y se hizo su representación sobre un gráfico a partir de una matriz de distancias euclidianas generada, respectivamente, de las coordenadas factoriales de ACP y AFC.

Análisis de clasificación

En un segundo paso, los caracteres continuos fueron parametrizados nuevamente en cinco clases, excepto el número de inflorescencias por nudo (PIN) que sólo ocurre en dos estados. Con los datos de los 36 individuos se realizó el Análisis de Clasificación (AC) y se calcularon el coeficiente de disimilitud de Sokal y Michener. Las divergencias entre accesiones se representaron en un dendrograma de árbol no-jerárquico utilizando el sistema de agregación del 'vecino más próximo' (Perrier y Jacquemound-Collet, 1999).

Es importante señalar que el coeficiente de Sokal y Michener utilizado otorga un peso similar a las concordancias y diferencias de modalidades entre individuos; por otra parte, el método de agregación del 'vecino más próximo' guarda los mismos principios que la clasificación jerárquica. Este último está muy difundido en estudios de clasificación, aunque difiere en dos puntos; primero, considera más importante definir dos elementos como próximos en la medida que difieren netamente de todos los demás, y segundo, realiza un ajuste de la disimilitud inicial a una distancia aditiva representada como un árbol y permite una representación más real de las divergencias. Este tipo de representación no es frecuente en la literatura genética, ya que muy pocos programas automatizados de análisis lo han incorporado, aunque es muy útil cuando es difícil de suponer un ancestro común de las accesiones en estudio, o no es de interés.

Resultados

El análisis de la variación permitió identificar los caracteres que presentan errores reducidos y considerar nueve cuantitativos para ACP y 12 cualitativos para AFC. El resto de caracteres entró al análisis como variables suplementarias. Para estimar las relaciones entre las accesiones y entre caracteres de los 32 individuos de los subgéneros *Tacsonia* y *Manicata* se utilizó un total de 21 de los 47 caracteres originales (Cuadro 4.3).

Cuadro 4.3. Descomposición de la variabilidad de los caracteres medidos en los subgéneros *Tacsonia* y *Manicata* que presentaron un alto valor de la relación genética/total y que fueron utilizados para los análisis factoriales.

Caracter ^a	Varianza			
	Genética	Error	Total	Genética/total
PLE	256.94	339.25	339.25	0.76
LWI	130.02	226.10	226.19	0.57
LLP	1476.20	2527.75	2527.75	0.58
LAL	461.90	887.37	887.37	0.52
DIN	0.05	0.08	0.08	0.65
BLE	477.33	613.57	613.57	0.78
FLE	664.96	1059.26	1059.26	0.63
FHL	562.59	895.46	895.46	0.63
FSP	105.58	135.87	135.87	0.78
TAP	1.12	0.82	1.94	0.58
SSH	1.04	0.54	1.58	0.66
SPA	1.05	0.64	1.69	0.62
LAS	2.88	2.39	5.27	0.55
LPP	1.83	1.17	3.00	0.61
LNP	1.40	0.85	2.25	0.62
BPE	0.52	0.43	0.95	0.55
BSH	1.57	0.74	2.31	0.68
BCO	1.39	1.07	2.46	0.57
FOR	0.75	0.52	1.27	0.59
FSF	3.23	0.95	4.18	0.77
FPP	1.90	0.84	2.74	0.69

a. En total fueron analizados 47 caracteres (Cuadro 4.1).

Las varianzas y las contribuciones de las variables activas en los primeros componentes del ACP, que comprenden al 68% de la variación total, aparecen en el Cuadro 4.4. Los individuos activos y suplementarios se representan en el primer plano factorial en la Figura 4.1a y las relaciones entre los caracteres principales y suplementarios en la Figura 4.1b. El primer componente (31% de variación) está correlacionado positivamente con la longitud del pecíolo y con la posición de los estigmas, y negativamente con el ancho de la hoja y la distancia de la invaginación de ésta a la inserción del pecíolo.

Las especies de los subgéneros *Tacsonia* y *Manicata* se distinguen de aquellas en los subgéneros *Passiflora*, *Decaloba* y *Psilanthus* porque la mayoría en estas últimas muestran

Cuadro 4.4. Varianza y cargas factoriales de los caracteres que más contribuyen en los tres primeros componentes principales en el análisis ACP en una colección de *Passiflora*.

	Factor 1		Factor 2		Factor 3	
Varianza (%)	31		21		16	
Varianza acumulada	31		52		68	
Cargas factoriales ^a	LLP	-0.47	LAL	0.36	BLE	-0.63
	LWI	-0.42	FLE	-0.64	DIN	-0.43
	PLE	0.37	FHL	-0.61	—	—
	FSP	0.38	—	—	—	—

a. El significado de las cargas aparece en el Cuadro 4.2.

hojas unilobular; por esta misma razón, *P. lanata* aparece cerca de este último grupo. Entre las especies del subgénero *Tacsonia*, las accesiones de *P. tripartita* var. *mollissima* de Colombia y Ecuador presentan peciolos y estigmas más largos. Entre las accesiones con estas características, *P. mixta* también presenta variación. Las accesiones de Boyacá y Valle, Colombia, tienen hojas más estrechas y una distancia más corta de la invaginación de la hoja a la inserción del peciolo. El segundo componente (21% de variación del total) refleja que los estigmas sobresalen más en *P. pinnatistipula* que en otras especies de subgénero *Tacsonia*. *Passiflora tripartita* var. *mollissima*, *P. tenerifensis*, *P. fimbriatistipula* y la accesión de *P. mixta* de Colombia central son similares, con flores e hipantios largos; mientras *P. tarminiana*, *P. cumbalensis*, y especialmente el grupo de *P. manicata* y *P. gracilens* se encuentran en el lado contrario del segundo componente. El tercer componente, que representa el 16% de la variabilidad total, muestra coeficientes altos asociados con valores igualmente altos para la longitud de las brácteas y el número de flores por nudo en las accesiones de *P. manicata*, *P. trinervia* y *P. tenerifensis*. Al extremo opuesto del componente aparecen *P. adulterina*, *P. crispolanata*, *P. cumbalensis* (de Cundinamarca, Colombia), *P. tripartita* var. *mollissima* y *P. trifoliata* que tienen sólo una flor por nudo y poseen brácteas menos marcadas.

En el Cuadro 4.5 se incluyen los coeficientes de los caracteres en los primeros cuatro factores del AFC (71% de variación total) y en las Figuras 4.2a y 4.2b se representan las posiciones de los individuos en los primeros dos factores. El primer factor se caracteriza por coeficientes altos para la forma de las estipulas, la presencia de antocianina en ellas, el ápice de las hojas y la orientación de las flores, que son los caracteres que más diferencian *P. tenerifensis*, *P. fimbriatistipula* y *P. lanata* de *P. mixta* y *P. tripartita* var. *mollissima*. En el grupo de *P. mixta* se observó una variación marcada en la orientación de flores; por ej., la accesión de Ecuador no tenía las flores erectas como las demás. El segundo factor está relacionado con características de la flor como el color de la faz adaxial de los pétalos y la presencia de pubescencia. Las accesiones de *P. tripartita* var. *mollissima* y *P. mixta* se presentaron como un grupo variable, debido principalmente a las diferencias en la pubescencia de flores, pero estos grupos se distinguieron de *P. tarminiana*, *P. cumbalensis*, *P. gracilens* y *P. trifoliata* por el color de la flor. *Passiflora x rosea* es menos pubescente que *P. pinnatistipula*. En este factor también se agrupan *P. trifoliata* y *P. gracilens* debido a la abundante pubescencia en las hojas y flores. El tercer factor, que no se incluye, se relaciona principalmente con la forma y color de brácteas y la posición de los nectarios en el peciolo. *Passiflora bracteosa* se coloca fuera del grupo de las *Tacsonia* clásicas y por su característica de brácteas libres igual que *P. fimbriatistipula* se aproxima a la accesión de *P. mixta* de Tolima (Colombia). Las accesiones de *P. manicata*, *P. cf. manicata*, *P. tenerifensis* y *P. mixta*

Cuadro 4.5. Caracteres que más contribuyen en los cuatro primeros componentes principales en el análisis AFC en una colección de *Passiflora*.

	Factor 1		Factor 2		Factor 3		Factor 4	
Varianza (%)	27		19		14		11	
Varianza acumulada	27		46		60		71	
Cargas factoriales ^a	SSH	0.6	LPP	0.33	BSH	0.56	TAP	0.26
	SPA	-0.28	FSP	-0.37	LNP	-0.32	BPE	-0.22
	LAS	-0.52	FPP	0.71	BCO	-0.31	—	—
	FOR	-0.23	—	—	—	—	—	—

a. El significado de las cargas aparece en el Cuadro 4.2.

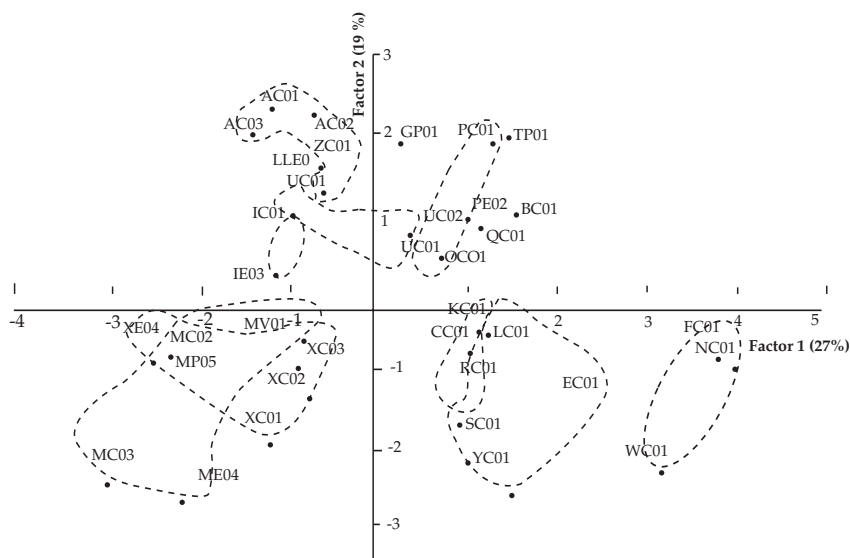


Figura 4.2a. Proyección de las accesiones de los subgéneros *Tacsonia*, *Psilanthus*, *Passiflora* y *Decaloba* sobre los dos primeros factores del AFC. Etiquetas de acuerdo con los datos en el Cuadro 4.1.

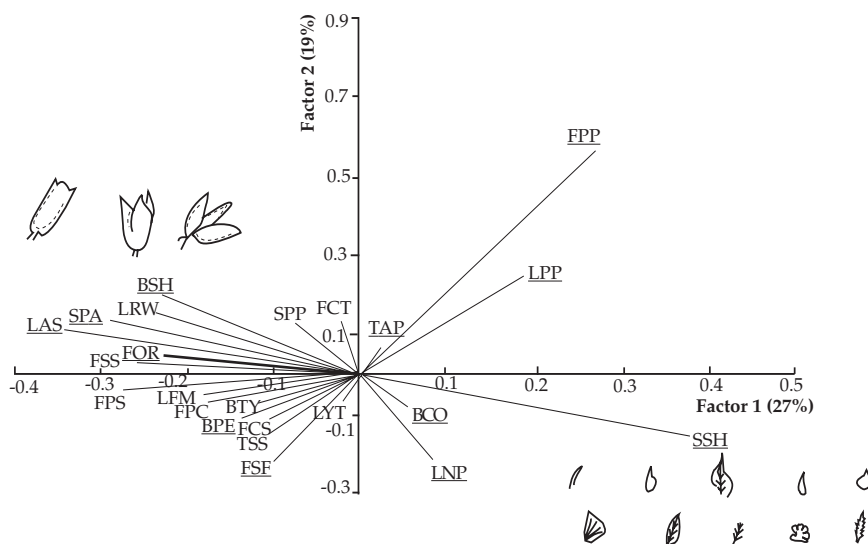


Figura 4.2b. Variables activas (subrayadas) y suplementarias proyectadas en el primer plano factorial del AFC. Los datos corresponden a 32 accesiones de los subgéneros *Tacsonia* y *Manicata*. Se observan las diferencias en la forma de las bracteas (BSH) y de las estípulas (SSH). Las etiquetas corresponden a los datos en el Cuadro 4.2.

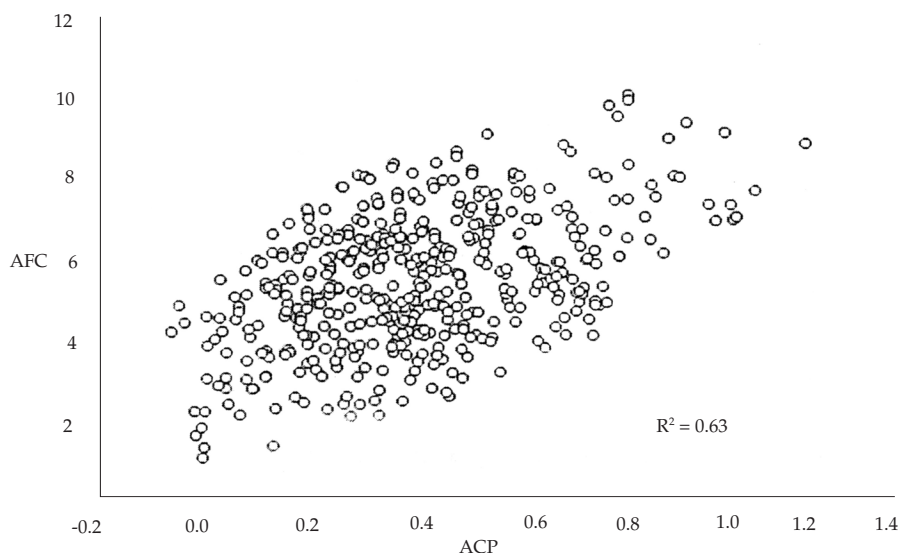


Figura 4.3. Correlación entre las matrices de distancias euclidianas calculadas a partir de las coordenadas factoriales del Análisis Factorial de Correspondencia (AFC) y del Análisis de Componentes Principales (ACP).

del Ecuador se localizan de manera dispersa. Finalmente, el cuarto factor, que al igual que el anterior no se incluye, separa las accesiones de *P. tarminiana*, la mayoría de las accesiones de *P. manicata*, *P. cf. manicata* y *P. antioquiensis* de las accesiones de *P. cumbalensis* gracias a las brácteas de tipo caducas presentes en las últimas. Killip (1938) utilizó este carácter para incluir a *P. manicata* y *P. antioquiensis* en el mismo subgénero *Granadillastrum* debido a las características de sus brácteas libres.

La matriz de correlación de distancias euclidianas del ACP y AFC (Figura 4.3) muestra la correspondencia estrecha que existe entre los conjuntos de datos. En una etapa siguiente fueron reparametrizadas las mediciones cuantitativas y se analizaron como cualitativas con el objeto de hacer una clasificación usando todos los parámetros.

Análisis de clasificación

El análisis de clasificación de las 36 accesiones con un total de 47 caracteres (incluyendo los 24 cuantitativos reparametrizados) reveló que el subgénero *Tacsonia* presenta marcadas diferencias en los niveles de divergencias morfológicas entre las especies y secciones (Cuadro 4.6 y Figura 4.4). Estos resultados coinciden con los obtenidos en el ACP y en el AFC y muestran que: (1) *Passiflora tripartita* var. *mollissima* y *P. mixta* presentaron una marcada proximidad; (2) un individuo de *P. mixta* de Ecuador se insertó entre los aglomerados de estas dos especies; (3) las accesiones de *P. cumbalensis* y *P. tarminiana* aparecen como dos grupos vecinos; (4) las cuatro especies forman un grupo compacto dentro del subgénero *Tacsonia*; y (5) los tres individuos de *P. manicata* forman un grupo separado, pero cercano, que incluye las accesiones de *P. cf. manicata*. Esto era de esperar ya que las hojas de *P. manicata* son trilobuladas con estípulas reniformes y las brácteas son ligeramente coalescentes a pesar de que el hipantio es más corto que en la mayoría de las especies de *Tacsonia*. Así, el subgénero *Manicata* se asemeja más a *Tacsonia* que a los dos otros subgéneros. Las similitudes de la flor de *P. manicata* y *P. pinnatistipula* se reflejan por su proximidad, siendo ligeramente más cercanas al grupo de especies de

Cuadro 4.6 Distancias medias de Sokal y Sheath entre especies de *Passiflora* (bajo la diagonal) e intra específicas de las principales especies (diagonal) a partir de las medidas morfológicas.

	pin	x ro	an	lan	cri	adu	ten	par	fim	bra	tri	sp.	cun	amp	tri	gra	mim	man	cf.	edu	lig	aln
<i>P. pinnatisipula</i>	0.36																					
<i>P. x rosea</i>	0.53																					
<i>P. antioquiensis</i>	0.52	0.53																				
<i>P. lanata</i>	0.67	0.68	0.58																			
<i>P. crispolanata</i>	0.59	0.60	0.64	0.66																		
<i>P. adlterina</i>	0.65	0.66	0.64	0.72	0.49																	
<i>P. tenerifensis</i>	0.61	0.65	0.60	0.57	0.61	0.67																
<i>P. paritiae</i>	0.64	0.66	0.63	0.72	0.61	0.67	0.66															
<i>P. finbriatisipula</i>	0.63	0.64	0.61	0.54	0.62	0.68	0.53	0.67														
<i>P. bracteosa</i>	0.68	0.69	0.67	0.75	0.64	0.70	0.70	0.56	0.71													
<i>P. triparita</i>	0.65	0.66	0.64	0.74	0.66	0.69	0.68	0.71	0.69	0.75	0.38											
<i>P. sp. 'Curuba India'</i>	0.66	0.70	0.64	0.74	0.66	0.72	0.68	0.71	0.70	0.75	0.66	0.53										
<i>P. cumbalensis</i>	0.63	0.64	0.62	0.71	0.64	0.70	0.66	0.69	0.67	0.73	0.60	0.64	0.45									
<i>P. ampullacea</i>	0.59	0.61	0.58	0.68	0.60	0.66	0.62	0.65	0.64	0.69	0.66	0.66.	0.63									
<i>P. trifoliata</i>	0.63	0.65	0.62	0.72	0.64	0.70	0.66	0.69	0.68	0.73	0.67	0.67	0.65	0.64								
<i>P. graciliana</i>	0.60	0.61	0.59	0.68	0.61	0.67	0.63	0.66	0.64	0.70	0.66	0.67	0.64	0.55	0.65							
<i>P. mixta</i>	0.62	0.63	0.61	0.71	0.63	0.69	0.65	0.68	0.66	0.72	0.49	0.63	0.63	0.63	0.64	0.63	0.45					
<i>P. manicata</i>	0.63	0.64	0.62	0.71	0.64	0.70	0.66	0.69	0.67	0.73	0.65	0.60	0.63	0.63	0.61	0.64	0.62	0.57				
<i>P. cf. manicata</i>	0.63	0.65	0.62	0.72	0.64	0.70	0.66	0.69	0.68	0.73	0.67	0.67	0.64	0.64	0.59	0.64	0.64.	0.57				
<i>P. edulis</i>	0.65	0.66	0.63	0.73	0.65	0.71	0.67	0.71	0.69	0.74	0.71	0.71	0.45	0.60	0.69	0.43	0.68	0.69	0.69			
<i>P. ligulata</i>	0.70	0.72	0.69	0.57	0.70	0.76	0.61	0.75	0.58	0.79	0.77	0.78	0.75	0.71	0.75	0.72	0.74	0.75	0.75	0.77		
<i>P. adnifolia</i>	0.67	0.68	0.65	0.75	0.67	0.73	0.69	0.72	0.71	0.76	0.73	0.73	0.71	0.62	0.71	0.45	0.70	0.71	0.71	0.35	0.78	
<i>P. trinervia</i>	0.73	0.75	0.72	0.53	0.73	0.79	0.64	0.78	0.61	0.82	0.80	0.81	0.78	0.75	0.79	0.75	0.71	0.78	0.79	0.80	0.79	0.82

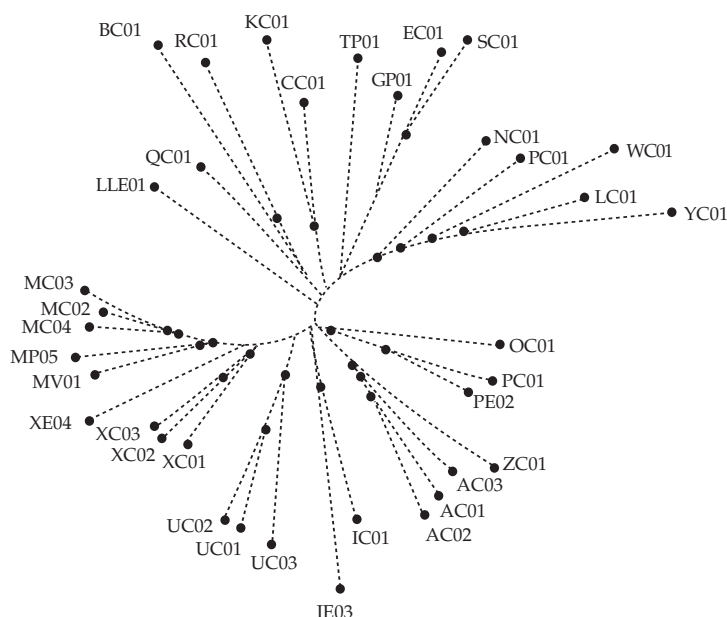


Figura 4.4. Representación radial de 35 accesiones de *Passiflora* obtenidas del análisis de clasificación (Neighbor-joining). Las accesiones están identificadas según el Cuadro 4.1.

Tacsonia comunes. *Passiflora pinnatistipula* es una especie de *Tacsonia* pero sus características particulares de corona y de estípulas la distinguen de aquellas más típicas como *P. tripartita* var. *mollissima*.

El grupo conformado por las accesiones restantes se divide en los grupos subalternos siguientes:

1. *Passiflora ampullacea* se localiza cerca a las especies comunes de *Tacsonia*, debido a la similitud de la hoja y las características de la bráctea.
2. *Passiflora bracteosa* y *P. parritae* se agrupan con *P. antioquiensis* gracias a la longitud de pedúnculo y forma de brácteas y hojas.
3. Las accesiones de *P. adulterina* y *P. crispolanata* forman una rama única, lo que confirma su proximidad taxonómica.
4. Las accesiones de *P. trifoliata*, *P. gracilens*, *P. edulis* y *P. alnifolia* se encuentran en la proximidad del tercer grupo según las características de la flor. *Passiflora edulis* y *P. alnifolia* no están clasificadas en el subgénero *Tacsonia*.
5. A pesar de que varias características cualitativas, como la forma de hojas, relacionan en un grupo las especies *P. tenerifensis*, *P. fimbriatistipula* y *P. lanata* con *P. ligularis* y *P. trinervia*, el parecido que resulta es discutible si se tiene en cuenta la afinidad filogenética, ya que las tres primeras poseen flores y frutos que son características de *Tacsonia*. Esto indica que los descriptores propuestos no son aplicables para *P. ligularis* y *P. trinervia*. Una situación similar se presenta con *P. gracilens* y *P. trifoliata* que se sitúan junto con las especies de los subgéneros *Passiflora* y *Decaloba*. Ambos casos

confirman la aplicabilidad de los descriptores para estudiar la variabilidad en las especies de los subgéneros *Tacsonia* y *Manicata*, pero muestran que son menos confiables en otros subgéneros de *Passiflora*.

Discusión

En este estudio se enfatizan las divergencias morfológicas de especies representativas de los subgéneros *Tacsonia* y *Manicata* y las relaciona con otras especies de los subgéneros *Passiflora*, *Decaloba* y *Psilanthus*. Las características morfológicas seleccionadas resultaron útiles para describir la variabilidad dentro y entre especies de los dos primeros subgéneros, pero se mostraron menos útiles para distinguir los cinco subgéneros de *Passiflora*. Para describir especies del género *Passiflora* se han utilizado varios criterios morfológicos, por ejemplo, Killip (1938) desarrolló una terminología que fue mejorada por Tillet (1988) y que sirvió como base para un artículo en el cual Jørgensen et al. (1984) publicaron un listado de las características mínimas necesarias que se deben tomar en cuenta cuando se realizan recolecciones de germoplasma de este género, pero no incluyeron las características de la hoja. La mayoría de caracteres recomendados por estos autores se refieren a las brácteas y las flores.

En este caso, donde se usaron 21 características derivadas de un total de 47, aparece que los 12 rasgos propuestos por Jørgensen et al. (1984) son necesarios, pero no suficientes, para la caracterización intraespecífica de especies del subgénero *Tacsonia*.

El ACP en este caso se hizo con las medidas cuantitativas y se encontró que los coeficientes más altos estuvieron relacionados con las características de las hojas y las dimensiones de la flor, la distancia del peciolo a la invaginación de los lóbulos, la posición de los estigmas y la longitud de las flores. Del esquema observado en la Figura 4.1a (el círculo de la correlación de caracteres cuantitativos) está claro que las características de la hoja presentan correlaciones entre ellas. Una relación similar se observa entre los rasgos de la flor, indicando que es posible una reducción de las dimensiones. Cuando las características cualitativas fueron analizadas y representadas en el primer plano factorial del AFC fue evidente una tendencia de relación de las variables por órganos. Caracteres como la pubescencia en las hojas y flores parecen relacionados, pero la mayoría de las observaciones en estas partes de la planta tienden a agruparse.

Los resultados de los análisis multivariados en este estudio señalan una estrecha relación entre *P. tripartita* var. *mollissima*, *P. mixta*, *P. cumbalensis* y *P. tarminiana*, constituyendo un subconjunto dentro de *Tacsonias*; no obstante, otros tipos de descriptores como los bioquímicos y moleculares podrían indicar la existencia o ausencia de un flujo de genes entre ellas. En términos de conservación, las accesiones de *P. tripartita* var. *mollissima* de Colombia y Perú son bastante similares y la accesión de *P. mixta* de Ecuador es una excepción en virtud de las características de las hojas y la flor. Un primer estudio morfológico de especies del subgénero *Tacsonia* (Villacis et al., 1998) también reveló las afinidades entre *P. mixta*, *P. tripartita* var. *mollissima*, *P. tarminiana* y *P. cumbalensis*. Este autor sugiere que el color de la corola no es un carácter distintivo para este taxón. El énfasis en el color rojo probablemente ha sido una causa de algunos errores de identificación de *P. mixta* en el pasado. Escobar (1980) comparó poblaciones de esta especie de Venezuela, Colombia y Ecuador con *P. cumbalensis* var. *goudotiana* y *P. cumbalensis* var. *cumbalensis* utilizando la forma de la bráctea, la pubescencia de la flor y

el tamaño de la hoja. Es interesante notar que las colecciones de *P. mixta* con flores glabras incluidas en el presente estudio también provienen del sur de Colombia y el norte de Ecuador, zonas donde existe un mayor contacto con *P. cumbalensis*, que es glabra. Las brácteas de estos especímenes también se asemejan más a las de *P. cumbalensis* que a las de los individuos pubescentes de *P. mixta*. El tamaño de la hoja varía a través del rango de distribución de *P. mixta* y en las áreas de contacto con *P. cumbalensis*, que tiene hojas pequeñas, se encuentran individuos glabros y pubescentes. Finalmente, Escobar (1980) concluye que una parte de la variación encontrada en *P. mixta* también puede ser explicada por la hibridación e introgresión con *P. cumbalensis* y otras especies menos abundantes, igualmente, parte de la variabilidad puede ser explicada por el aislamiento geográfico de poblaciones pequeñas y por la erosión genética.

Los resultados de la morfología de las especies en este estudio contrastan con la situación taxonómica actualmente aceptada de *P. manicata* y propuesta por Escobar (1988), en la cual esta especie se coloca fuera de subgénero *Tacsonia*. Así, *P. manicata* se agrupa con las especies del subgénero *Tacsonia*, con una similitud próxima hacia *P. pinnatistipula*. *Passiflora tripartita* var. *mollissima* es simpátrica con *P. pinnatistipula* en el nororiente colombiano, donde fue recolectado *Passiflora* x *rosea*, su híbrido putativo. Este híbrido tiene una flor más pequeña que *P. tripartita* var. *mollissima*, estipulas y hojas más grandes que *P. pinnatistipula*, pero en el dendrograma *Passiflora* x *rosea* se agrupa con las accesiones de *P. pinnatistipula* debido a su marcada similitud morfológica global.

Se encontró una estrecha relación entre *P. antioquiensis* (sección Colombiana), *P. parritae* (sección Parritana) y *P. bracteosa* (sección Anómala), pero *P. tenerifensis* aparece como no relacionada con *P. antioquiensis*, aunque son clasificadas en la misma serie *Leptomischae* de la sección Colombiana. De las especies de esta sección, sólo *P. crispolanata* y *P. adulterina* muestran una afinidad próxima. En los resultados del presente estudio, *P. lanata*, *P. tenerifensis* y *P. fimbriatistipula* (sección Fimbriatistipula) parecen relacionadas por las similitudes de las características de las flores, hojas, brácteas y estipulas, pero en la taxonomía actual la última especie se excluye de la sección Colombiana. La similitud de las distribuciones geográficas entre *P. antioquiensis* y *P. parritae*, y entre *P. tenerifensis* y *P. fimbriatistipula* es paralela con algunas proximidades morfológicas. En consecuencia, la división actual de sección Colombiana en las series *Leptomischae* y *Colombianae*, y la estructura de las secciones Colombiana, Fimbriatistipula, Parritana, Boliviana y Trifoliata, propuesta por Escobar (1988), no es concordante con este estudio morfológico. La inclusión de más accesiones por especies y el uso de herramientas moleculares podría clarificar, aún más, la situación actual. Sería, también, interesante enfatizar en las afinidades entre *P. mixta*, *P. cumbalensis*, *P. tarminiana* y *P. tripartita* en Ecuador y Perú.

En las especies de los subgéneros *Passiflora* (*P. ligularis* y *P. edulis* f. *flavicarpa*) se observó una marcada heterogeneidad debida, principalmente, a las diferencias foliares. *Passiflora edulis* f. *flavicarpa* se agrupó con *P. alnifolia* (subgénero *Decaloba*) y *P. ligularis* con *P. trinervia* (subgénero *Psilanthus*). Estos resultados indican que los descriptores seleccionados para el estudio de la variación entre especies del subgénero *Tacsonia* no son adecuados para describir las divergencias entre especies de los otros subgéneros. Esto no es sorprendente ya que caracteres como los 'palli', los 'radii' y otras formas de la hoja ausentes en las especies de subgénero *Tacsonia* no fueron incluidos en la lista de descriptores de las especies de este subgénero.

Conclusiones

El estudio de la variación morfológica entre las especies del subgénero *Tacsonia* es posible con la lista de descriptores usada en el presente estudio, pero también está claro que es necesario aumentarlos para describir mejor la variación a nivel intra-accesiones. Treinta, de las 36 accesiones en el estudio, fueron también analizadas por las técnicas de AFLP (Segura et al., 2001) y se confirmaron las similitudes entre *P. tripartita* var. *mollissima*, *P. mixta*, *P. tarminiana* y *P. cumbalensis*. La separación de este grupo de especies de otras secciones también fue evidente. Con el uso de estos marcadores moleculares, la posición de especies de otros subgéneros, inclusive *Manicata*, parece estar más de acuerdo con la taxonomía actual.

Los equipos de trabajo en recursos genéticos deben estudiar las posibilidades del uso de un listado más resumido de descriptores morfológicos en las colecciones de *Passiflora* de los países andinos. En este trabajo, de un listado de 47 descriptores fueron seleccionados nueve cuantitativos y 12 cualitativos para la caracterización de especies del género *Passiflora*, ya que estos producían una mejor descripción de la variación entre especies de los subgéneros *Tacsonia* y *Manicata*. Las características de flor y de hoja fueron las más descriptivas, mientras que las de las brácteas y las estipulas proporcionaron información convergente con la taxonomía actual.

El análisis resaltó la estrecha similitud entre las especies más comunes de los subgéneros *P. mixta* y *P. tripartita* var. *mollissima*. *Passiflora cumbalensis* y el cultivar recientemente descrito como especie *P. tarminiana* aparecen estrechamente relacionados con las especies comunes. En contraste, especies de las secciones Colombiana, Parritana, Fimbriatistipula, Anómala y Poggendorffia aparecen claramente separadas de las especies comunes. *Passiflora manicata* y *P. pinnatistipula* formaron grupos intermedios entre el grupo común y las cinco secciones antes mencionadas. Como era de esperar, *Passiflora x rosea* aparece relacionada con *P. pinnatistipula*.

El análisis basado en los datos morfológicos tiene una aplicación inmediata. Los rasgos morfológicos, a veces, son susceptibles de modificaciones debidas al ambiente, pero este hecho no elimina la posibilidad de detectar descriptores morfológicos estables que puedan ser utilizados para medir la divergencia entre especies. En los países andinos las herramientas moleculares no están todavía extensamente disponibles para investigadores que trabajan con recursos genéticos de *Passiflora*, por tanto, es posible el uso de caracteres morfológicos para caracterizar y clasificar las accesiones en grupos similares con el fin de planificar el trabajo de mejoramiento y obtener información básica para trazar estrategias de conservación in situ y ex situ. No obstante, se recomienda realizar estudios de biología floral para clarificar el complejo de especies alrededor de *P. tripartita* var. *mollissima* y las divergencias entre especies del subgénero *Tacsonia*, usando marcadores moleculares como complemento a la descripción morfológica.

Referencias

- Escobar, L. A. 1980. Interrelationships of the edible species of *Passiflora* centering around *Passiflora mollissima* (H.B.K.) Bailey, subgenus *Tacsonia*. Ph.D. Thesis, University of Texas. p. 646.
- Escobar, L. A. 1985. Biología reproductiva de *Passiflora manicata* e hibridación con la
-

- curuba, *Passiflora mollissima*. Actualidades Biológicas 14(54):111-121.
- Escobar, L. A. 1988. Passifloraceae. En: Pinto, P. y Lozan, G. (eds.). Flora de Colombia. Universidad Nacional de Colombia, Bogotá, Colombia. 135 p.
- Jørgensen, P. M.; Lawesson, J. E., y Holm-Nielsen, L. B. 1984. A guide to collecting passion flowers. Ann. Missouri Bot. Gard. 71:1172-1174.
- Holm-Nielsen, L.; Jørgensen, P.M.; y Lawesson, J. E. 1988. Passifloraceae. En: Harling, G. y Andersson, L. (eds.). Flora of Ecuador. University of Göteborg, Copenague. 31:130 p.
- Killip, E. P. 1938. The American species of Passifloraceae. Field Museum of Natural History Botanical. Botanical. 2. 613 p.
- Lobo, M. 1997. Informe final proyecto pasifloras, Corporación Colombiana de Investigación Agropecuaria (Corpoica). Informe final. Cooperación técnica IPGRI-BID No. ATN/SF-44356-RG. Diversidad, Conservación y Uso Sostenible de los Recursos Genéticos de Frutales Nativos de América Tropical. Mayo, 1997, Corpoica, Rionegro, Colombia.
- Perrier, X. y Jacquemoud-Collect, J. P. 1999. DARWIN. Version 3.0. CIRAD-FLHOR. Montpellier.
- Schoëninger, G. 1986. La Curuba. Técnicas para el mejoramiento de su cultivo. Editora Guadalupe Ltda., Bogotá, 255 p.
- Tillett, S. S. 1988. Passionis passifloris II. En: Ernstia, (ed.). Terminología. Herb. Fac. Agron. Univ. Central de Venezuela, Maracay. p. 1-49.
- Villacis, L. A.; Vega, J.; Grum, M.; y Coppens d'Eeckenbrugge, G. 1998. Morphological characterization of Andean Passifloras (*Passiflora* spp.) from Ecuador. Plant Gen. Res. Newsl. 115: 51-55.
-

Conceptos y Mediciones Útiles para la Caracterización de Germoplasma

A continuación se presentan algunos conceptos y mediciones generales relacionados con la temática de los estudios de caso presentados en este boletín. Con ellos se quiere recopilar las experiencias adquiridas a través de los trabajos con diferentes colecciones de germoplasma de especies cultivadas y ofrecer una visión adicional de la práctica de la caracterización en la cual existen tantas variantes como especies cultivadas. El enfoque gira alrededor de los objetivos que se buscan en la caracterización de los recursos genéticos, en un intento por proporcionar criterios adicionales a los mejoradores, curadores y responsables de las colecciones en los bancos de germoplasma.

Análisis estadístico de datos de caracterización morfológica

*J. L. Chávez**

Uno de los aspectos esenciales en los trabajos de caracterización de especies es la descripción de una serie de colecciones desde el punto de vista de sus atributos morfológicos. Para medir la variabilidad genética o generar algunos estimadores de ella se deben conocer el origen geográfico de la colección y las fuentes de diversidad. Con la caracterización morfológica se espera conocer el nivel de variabilidad de la colección, la estructura genética de cada una de las poblaciones (accesiones) y las herramientas metodológicas más adecuadas, además, los análisis estadísticos y estimadores más adecuados que responden mejor al objetivo de la caracterización.

Medición de la variabilidad genética

Para medir la variabilidad genética de una colección se utilizan, básicamente, las evaluaciones morfoagronómicas, la caracterización mediante marcadores basados en proteínas o isoenzimas y fragmentos de ADN y en el número o la riqueza de cultivares (variedades locales, mejoradas e introducidas) en un sitio.

Medición de la variabilidad mediante evaluaciones morfoagronómica. Para la caracterización y la evaluación morfoagronómica es necesario sembrar el material genético y evaluarlo en lotes experimentales o en campos de agricultores. En esta etapa, además, se hace una valoración agronómica del potencial productivo y se evalúan la tolerancia a plagas y enfermedades y a estrés bajo condiciones abióticas, en un diseño experimental con testigos de referencia (variedades comerciales o cultivares de uso común en la región).

Cuando la variabilidad de una colección se cuantifica en términos genéticos, las caracterizaciones y evaluaciones se deben hacer en varias localidades y a través de estaciones de crecimiento. Una vez que se reúne la información se procede a su análisis estadístico mediante los métodos multivariados de descripción, ordenación o clasificación

* Especialista en conservación de recursos fitogenéticos in situ. IPGRI-CODESU, Centro de Investigación y de Estudios Avanzados (IPN), México.

adecuados. Sin embargo, los análisis de varianza clásicos son los de mayor utilidad para separar los efectos genéticos de los ambientales, es decir, para evaluar la variabilidad en función de los caracteres morfológicos altamente heredables. Los ensayos de caracterización en una localidad no permiten separar el efecto ambiental del genético. Goodman y Paterniani (1969) proponen la fórmula que aparece al pie de página³ para estimar la repetibilidad del carácter morfológico, en otras palabras, la heredabilidad (γ), donde $\gamma \geq 1$ en un carácter significa estabilidad a través de localidades o ambientes, bien que sea que se utilice esta fórmula o el análisis de varianza al clásico, las estimaciones se basan en los componentes de la varianza genética (varianza de las colecciones entre la varianza fenotípica).

Medición de la variabilidad basada en proteínas de semillas, isoenzimas y ADN. La biología molecular ha contribuido de manera significativa para la mejor descripción de la variabilidad genética de las colecciones y su filogenia. Los marcadores moleculares son herramientas útiles para medir la variabilidad genética; por ej., las isoenzimas o las proteínas de semillas describen, aunque con pocos alelos polimórficos, la diversidad genética de las colecciones. Aquí es necesario tener en cuenta que las bandas polimórficas son las expresiones de los genes (aloenzimas) separadas en un campo eléctrico. En el caso de los marcadores basados en ADN (RFLP's, RAPD's, AFLP's, SSR's o Microsatélites) o secuencias de ADN estas bandas tienen una mayor objetividad debido a que son secuencias de ADN fragmentado.

La aplicación de uno u otro marcador depende de la capacidad y de la infraestructura disponibles en el laboratorio. Cada marcador ofrece diferentes mediciones, las enzimas o proteínas definen locus aloenzimáticos con dos o más alelos. Los marcadores de ADN definen diferentes bandas polimórficas o loci polimórficos. Los RFLP's y SSR's establecen estimaciones indirectas de la divergencia de nucleótidos debido a su fragmentación específica del ADN y su expresión codominante. Los RAPD's son marcadores dominantes, lo que implica ciertas complicaciones en los análisis de la información (los dominantes están presentes y la forma recesiva está ausente) y una sobrestimación del nivel de diversidad. Para el análisis de datos moleculares se sugiere consultar el artículo de Ayad et al. (1995).

Con los datos moleculares de cada una de las entradas que se describen en la colección se procede a la estimación de la diversidad genética mediante la riqueza alelica, el coeficiente de diversidad de Nei (índice de diversidad), el nivel de heterocigosidad (observado y esperado), el índice de fijación, el grado de desequilibrio y el porcentaje de polimorfismo, entre otros.

Medición mediante el número de cultivares en un sitio. Este es un enfoque reciente donde se adoptan los conceptos de diversidad de especies según los análisis de biodiversidad ecológica o biológica. Para esta medición, las accesiones o variedades locales provenientes de una región particular se clasifican en una serie de clases o rangos definidos en función de la expresión fenotípica de los caracteres morfológicos. Por

3. Heredabilidad: $\gamma = \frac{\sigma_g^2}{\sigma_a^2 + \sigma_{ga}^2}$. Los valores $\gamma > 1$ indican que la medida entre accesiones tiene más expresión genética que efecto del ambiente e interacciones. Si $\gamma < 1$, la heredabilidad es baja y existe efecto del ambiente sobre la expresión del carácter (Goodman y Paterniani, 1969).

ejemplo, Louette et al. (1997) y Aguirre et al. (2000) estimaron los índices de diversidad de Shannon, Simpson y Margalef a partir de la clasificación de una serie de muestras de maíz por clase (raza, mezcla racial, variedad criolla, variedad mejorada, blanco, negro, entre otros) y ambiente (comunidad o localidad). Una vez definidas las clases, se clasifican todas las accesiones y se estiman los índices de diversidad con el objeto de determinar la variabilidad genética preservada por las comunidades de agricultores en función de la riqueza de morfotipos, varietal, o bien, de la riqueza de cultivares en una área determinada.

Representatividad de la colección

En el sentido teórico la representatividad de la variabilidad de una colección es una función del universo de muestreo o de la variabilidad total de la especie. En términos prácticos esto significa que es difícil definir la representatividad de una región ya que no se conoce la diversidad total y sólo se obtienen algunas estimaciones. Entre algunas de las complejidades se encuentran la alta diversidad de formas y propósitos de la cuantificación de la variabilidad incluyendo las metodologías, las diferentes formas y tamaños de las muestras que originaron las accesiones, los antecedentes documentados de la variabilidad de la muestra y el desconocimiento mismo de la diversidad objeto de estudio.

Mediante la teoría de probabilidades se pueden obtener algunos estimadores y la representatividad puede ser definida de las formas siguientes: (1) la colección representa el rango completo de variación en todo el universo de la muestra, y (2) el patrón de variación de la colección se asemeja a la variación total existente de la especie o de la región en la muestra. En otras palabras, cada entrada o accesión representa al menos una réplica de un rango de la variación total, o bien, los estimadores de la variación genética son aleatorios y de mínima varianza. Sin embargo, todos estos conceptos son poco aplicados para el caso de alelos raros o menos frecuentes que son ejemplos, teóricamente, únicos de la variación genética. Los conceptos anteriores de representatividad están relacionados directamente con la exploración y recolección del germoplasma que constituye la colección bajo estudio. –Para una mayor información sobre muestreo se recomienda consultar los trabajos de Brown y Briggs (1991)–. Cuando se dispone de datos previos es posible verificar o medir la representatividad de la variabilidad genética de una colección en relación con la diversidad total o esperada, bajo este aspecto se incluyen los ejemplos de erosión genética o de aparición de nuevas combinaciones genéticas. Los estimadores comúnmente usados para este fin son el índice de fijación (F_{st}), el grado de desequilibrio de ligamiento (D) y el grado de divergencia entre las poblaciones (F_{st} o G_{st}), que comparan la diversidad de la colección bajo estudio con la diversidad previamente estimada en colecciones similares o del mismo origen geográfico. Para obtener todos estos estimadores se utilizan datos de caracterización moleculares.

Investigación sobre la estructura genética

El análisis molecular (isoenzimas, proteínas o secuencias de ADN) es la técnica más reciente para conocer la estructura genética de cada una de las entradas de una colección de germoplasma. Con la información resultante de los loci identificados se procede a la descripción de la cada colección en función de su nivel de polimorfismo, la riqueza alélica,

el nivel de heterocigosidad observada y esperada, la frecuencia de alelos raros en cada accesión, el número de alelos por locus polimórfico y el índice Nei de diversidad genética.

Una vez que se conoce la estructura genética de cada entrada o accesión se procede a la estimación de las relaciones genéticas o de diversidad entre y dentro de las colecciones o accesiones. Adicionalmente, la biología reproductiva y la forma de reproducción desempeñan una función importante en la diversidad misma del germoplasma. Hamrick y Goldt (1990) proporcionan información de estas diferencias y similitudes mediante datos isoenzimáticos. La recopilación de una cantidad mayor de datos en este sentido permitirá respaldar la hipótesis de que las especies de polinización cruzada presentan mayor variabilidad genética que las especies de autofecundación y propagación vegetativa. Hacia el interior de cada colección se pueden establecer las diferencias o similitudes entre sus entradas mediante los índices de proximidad respectivos, utilizando las pruebas estadísticas pertinentes como ji-cuadrada (χ^2) y otras para establecer las posibles relaciones genéticas y la cuantificación del grado de divergencia entre las poblaciones o accesiones (F_{st} o G_{st}). Una de las formas más comunes de presentar gráficamente los resultados de este análisis consiste en realizar un ordenamiento o clasificación mediante el análisis de conglomerados de agrupación jerárquica.

De la misma manera, cuando se conoce la estructura genética de las accesiones de una colección es posible identificar las relaciones filogenéticas entre las poblaciones provenientes de diferente o del mismo origen geográfico, lo que permite conocer el patrón de distribución geográfica de la variabilidad genética y sus orígenes.

Identificación de duplicados

La detección de duplicados en una colección de germoplasma es tema de discusión en la conservación ex situ. La identificación propia de duplicados está en función de la valoración de los caracteres genéticos de una población o accesión. El germoplasma se preserva para su utilización posterior como una concentración de valores utilitarios, o bien, en función de su importancia como recurso genético, alimentario-nutricional, industrial, artesanal, medicinal, ceremonial y religioso, entre otros. Con esta premisa se pueden determinar los duplicados. En la situación más simple si un material **A** tiene un conjunto de valores o características deseables **Y** y, a su vez, una accesión **B** tiene un conjunto de valores **X**, de tal forma que **Y** y **X** son estadísticamente equivalentes, similares o iguales, entonces, se asume que **A** = **B**, por tanto, se estaría señalando que **A** y **B** son duplicados.

Para establecer los elementos comparativos y definir la situación de duplicados se deben realizar todas las valoraciones necesarias que deben estar en función de los objetivos de la preservación de los recursos genéticos; por ej., mediante una evaluación agronómica dos materiales son similares si responden estadísticamente igual en rendimiento y comportamiento a través de ambientes. Por otro lado, en un análisis SSR's se puede hacer una comparación similar y si todos los loci son significativamente iguales entre dos poblaciones o colecciones, se considera que su índice de divergencia (F_{st} o G_{st}) no es significativo.

Dependiendo de la precisión o validez de las comparaciones, en sentido estricto de la genética de poblaciones se pueden encontrar ciertas similitudes o 'igualdades' entre dos accesiones aunque no sean exactamente duplicados. La estimación de la estructura genética de una población o de una colección de germoplasma es el resultado de las

herramientas de campo o laboratorio utilizadas y del comportamiento de los testigos seleccionados para obtener los estimadores. La estructura genética de una población es comparable con otra si los protocolos y estimadores utilizados son los mismos o similares para ambas.

Una de las aplicaciones más difundida es la detección de duplicados o violación a los derechos de 'propiedad intelectual de las variedades comerciales'. Los microsatélites o SSR's suelen ser eficientes para comprobar o dilucidar posibles violaciones de propiedad. La separación en un campo eléctrico de los diferentes segmentos de ADN permite las comparaciones entre las poblaciones, ya que las porciones o segmentos extremadamente variables en su longitud y peso son característicos o específicos de cada población de plantas. Por tanto, la comparación y detección de posibles duplicados o variedades se hace con cierta facilidad, sin menospreciar los caracteres morfológicos y, más bien, complementando las señas de identificación mediante fragmentos de ADN que son independientes del estado de desarrollo de la planta y del ambiente en que ésta crece.

Detección de genes especiales o alelos particulares

La genética de poblaciones se puede considerar como la detección de poblaciones variantes raras o especiales. La utilidad de estas variantes o genes especiales y la posibilidad de encontrarlos depende, previamente, de la estrategia de muestreo, siendo su presencia una de las grandes riquezas de los recursos genéticos (Marshall y Brown, 1975). La detección de genes de resistencia a estrés biótico, abiótico y adaptabilidad a condiciones extremas es de primordial importancia. Sin embargo, las variantes raras también pueden ser irrelevantes en casos particulares o no tener connotación estratégica importante. El conocimiento a priori o la búsqueda de la variante deseable reviste, en la mayoría de los casos, más importancia en la exploración, recolección y valoración genética de una colección.

En las colecciones de germoplasma, por ejemplo, es común detectar un arreglo específico de genes que confieren resistencia a un patotipo de enfermedades y que se encuentran reunidos en cierto grupo de materiales provenientes de la región donde prevalecen las condiciones ambientales óptimas para el desarrollo de dichas enfermedades.

Las variabilidades fenotípica y genética de una colección de germoplasma pueden ser medida utilizando varios coeficientes e índices, entre ellos:

Variabilidad de caracteres morfológicos

- Coeficiente de variación genética
- Varianza genética (entre y dentro de acciones)
- Heredabilidad o repetibilidad de un carácter - γ
- Índice de Shannon-Weber (clases por caracteres con escala nominal).

Variabilidad por el número de cultivares en un sitio

- Índice de Shannon-Weber
- Índice de Simpson
- Índice de Margalef
- Índice de Berger-Parker.

Variabilidad a nivel genético

- Riqueza alélica -A
 - Coeficiente o índice de diversidad genética (índice de Nei) -h
-

- Heterocigosidad (esperada y observada) -H
- Índice de fijación -F
- Grado de desequilibrio de ligamiento, esencial en especies alógamas -D
- Grado de divergencia entre las poblaciones.

Referencias

- Aguirre, A.; Bellon, M.; y Smale, M. 2000. A regional analysis of maize biological diversity in Southeastern Guanajuato, México. *Econ. Bot.* 54:60-72.
- Ayad, W.; Hodgkin, T.; Jaradat, A.; y Rao, V. R. (eds.). 1995. Molecular genetic techniques for plant genetic resources. Report of an IPGRI Workshop 9-11, October 1995. Roma, Italia.
- Brown, A. H. y Briggs, J. D. 1991. Sampling strategies for genetic variation in ex situ collections of endangered plant species. En: Falk, D.A. y Holsinger, K. E. (eds.). *Conservation of rare plants: Biology and genetics*. Oxford, Reino Unido. Oxford University Press.
- Goodman, M. M. y Paterniani, E. 1969. The races of maize III. Choice of appropriate characters for racial classification. *Econ. Bot.* 23:65-273.
- Hamrick, J. L. y Goldt, M. J. 1990. Allozyme diversity in plant species. En: Brown, H. A.; Clegg, M.; Kalher, A.; y Weir, B. (eds.). *Plant population, breeding, and genetic resources*. Sinauer Associates Inc. Mass. E.U.
- Louette, D.; Charrier, A.; y Berthaud, J. 1997. In situ conservation of maize in México: Genetic diversity and maize seed management in a traditional community. *Econ. Bot.* 51:20-38.
- Louette, D. y Marshall. 1995. A basic sampling strategy: Theory and practice. En: Guarino, L.; Rao, V. R.; y Reid, R. (eds.). *Collecting plant genetic diversity: Technical guidelines*. CAB-International, IPGRI, FAO, IUCN, UNEP. p. 75-91.
- Marshall, D. R. y Brown, A. H. 1975. Optimum sampling strategies in genetic conservation. En: Frankel, O. H. y Howkes, J. G. (eds.). *Genetic resources for today and tomorrow*. Cambridge, Reino Unido, Cambridge University Press.

Caracterización de germoplasma

*Gustavo Ligarreto**

Variabilidad genética de colecciones

La medición de los caracteres cualitativos y cuantitativos de alta heredabilidad, o que se transmiten a la descendencia del germoplasma en cualquier ambiente, se conoce como caracterización y permite determinar el grado de similitud entre las accesiones por medio de su apariencia morfológica o fenotipo y de variabilidad en la colección. Esta variabilidad se mide con pocas o muchas variables o descriptores cuyos datos conforman una dispersión de puntos con una dirección o vector e interrelacionan para conformar las distancias genéticas entre las accesiones. Estas distancias, a su vez, se pueden graficar de diferentes formas, siendo los dendrogramas y la dispersión de puntos en un plano cartesiano las de más fácil interpretación.

* Profesor Asociado. Universidad Nacional de Colombia, Facultad de Agronomía, Bogotá.

El conocimiento que posee el investigador o curador sobre la colección en estudio es importante para decidir cuáles son los grupos de variabilidad de germoplasma a través de un diagrama o conjunto de distancias, es decir, le permiten analizar y discutir por qué a partir de los datos es posible inferir la similitud de las accesiones. También son importantes los datos de pasaporte ya que, en ocasiones, permiten inferir sobre el tipo de evolución, domesticación, adaptación y características de selección sobre los materiales.

Es importante conocer el estado de arte o trabajos previos sobre descriptores y métodos de análisis de datos de la especie, debido a que la toma de muchas variables en un alto número de accesiones y plantas por accesión hace costoso el ejercicio de la caracterización y dificulta su análisis. En lo posible se deben usar pocos descriptores de alta heredabilidad con el objeto que la variabilidad resulte suficientemente discriminada. Para el análisis existe un alto número de distancias y coeficientes de similitud, sin embargo, distancias como la euclidiana es de fácil análisis e interpretación y es muy útil cuando se tienen variables cuantitativas. Para los análisis mixtos de variables cualitativas y cuantitativas se debe preferir la distancia métrica de Gower.

Cuando se tienen colecciones con muchas accesiones que dificultan el proceso de caracterización, se puede hacer una descripción de tipo evaluativo con un número reducido de variables (cinco o seis) para formar grupos grandes que posteriormente son caracterizados por separado relacionando sus bases de datos. Otra forma consiste en tomar al azar aproximadamente el 10% de las accesiones de cada grupo o de la colección total y luego extrapolar la información a esta última. Si el número de accesiones no es muy grande y se tienen recursos económicos suficientes lo ideal es caracterizar toda la colección. También es importante mencionar que la caracterización de todas las colecciones existentes en un banco de germoplasma puede ser un esfuerzo grande que no siempre es utilizado por los investigadores o productores. Es interesante conocer cuáles colecciones demandan los investigadores y las instituciones para invertir en su caracterización buscando, con ello, un mayor retorno a la inversión de recursos en el conocimiento del germoplasma.

Estructura genética

La estructura genética de la población es el resultado de la interacción de factores de selección, mutación génica, migración y recombinación que pueden reducir o aumentar los niveles de variación en ella. La reducción de la diversidad genética de una especie produce pérdida de la variación necesaria para el mejoramiento, por consiguiente, con el mantenimiento del germoplasma se persigue evitar las alteraciones de la estructura genética de las accesiones.

En el mantenimiento del germoplasma no se ha prestado la suficiente atención a la constitución genética de las colecciones conservadas. Como ejemplos se pueden mencionar: (1) el efecto de la selección natural que ocurre al aumentar ciertos genotipos a través de sucesivas generaciones; y (2) los cambios aleatorios en las frecuencias de los genes a través de las generaciones, en especial los causados por los errores de muestreo. Existe abundante información sobre el mantenimiento de germoplasma *ex situ*, no obstante, aún existen fallas en los tamaños de muestra poblacional y en los procedimientos de recolección y de regeneración, que son importantes para mantener la estructura genética de la población.

Detección de duplicados

Mediante la identificación taxonómica y caracterización morfológica de todas las accesiones de una colección es posible identificar copias de un mismo genotipo o cultivar. Estas copias son el producto de varias recolecciones de un mismo material mantenido por diferentes agricultores en una misma región geográfica, o de un mismo cultivar que ha sido llevado a diferentes regiones bajo nombres diferentes, también puede ser el resultado de donaciones de germoplasma o de la recolección de numerosas muestras de cultivares que crecen en áreas geográficas muy extensas.

El material duplicado puede ser identificado a través de la aplicación de descriptores morfológicos, bioquímicos y moleculares, entre otros. Los descriptores de alta heredabilidad, entre los marcadores morfológicos, son los que más contribuyen en la detección de duplicados ya que tienen poco efecto del ambiente sobre la expresión de la característica evaluada, es decir, son variables que discriminan bien las accesiones. Los marcadores bioquímicos y moleculares son una herramienta potencial para la identificación efectiva y eficiente de material duplicado, siendo menos costosos los primeros.

La detección de duplicados se puede utilizar para: (1) reducir las colecciones a tamaños manejables, identificar variedades, monitorear la variabilidad genética y determinar la distribución geográfica de las especies. En este caso, el método para detectar duplicados consiste en la caracterización morfológica inicial y en el procesamiento de los datos para conformar grupos de accesiones similares, que posteriormente se siembran en el campo con el fin de comparar las accesiones dentro de grupos en todas las características morfológicas seleccionadas; y (2) comparar, en una segunda etapa, las accesiones similares por morfología mediante patrones electroforéticos de isoenzimas, proteínas y marcadores moleculares. Una vez detectados los duplicados, se selecciona la entrada más representativa dentro del grupo para mantenerla en la colección.

Divergencias morfológicas interespecíficas del subgénero *Tacsonia* (*Passiflora*)

*Sergio Segura**

En este resumen se explican conceptos básicos del análisis de la variabilidad vegetal y las técnicas estadísticas asociadas con un ejemplo aplicado a las pasifloras andinas. Se describe una posibilidad de análisis utilizando descriptores morfológicos cuando el objetivo del estudio es la revisión de las divergencias entre especies cultivadas y silvestres relacionadas.

En la primera parte se citan los marcadores de diversidad diferentes y más comunes, haciendo énfasis en los descriptores morfológicos. Aunque son conceptos muy conocidos, es importante señalar el valor informativo de este tipo de descriptores, tomando en cuenta la etapa de investigación genética en que se encuentra la mayor parte de los grupos de especies vegetales en los países de la región.

* Diversidad Genética de Frutales. Universidad Autónoma, Chapingo, México.

La metodología utilizada con las pasifloras andinas en el Caso 4 de este boletín es un ejemplo donde las características morfológicas fueron depuradas antes de utilizar métodos factoriales y de clasificación para sintetizar las relaciones entre variables y entre accesiones. Con otros grupos de plantas es posible seguir este camino de manera total o parcial, dependiendo del objetivo que se persiga.

Los marcadores de la diversidad en plantas

Los seres vivos individuales forman la unidad básica para el estudio de la variabilidad biológica, pero el nivel de observación de esta variabilidad se traduce en disciplinas diferentes del estudio de la diversidad natural. El estudio de las diversidades ecológica, específica y genética se diferencian según su escala de organización en ecosistemas, especies y poblaciones o individuos. La comparación de la variabilidad biológica genética descansa sobre el concepto de descriptor. Un descriptor es una característica que puede ser observada y/o medida por algún medio a partir de los individuos. Una clasificación muy generalizada divide los descriptores en plantas por la herramienta utilizada para su observación. Así se tienen descriptores morfológicos, bioquímicos y moleculares, aunque es de interés actual agregar a estas informaciones otro tipo de características como las etnobotánicas y climáticas (Perrier, 1998).

En la mayoría de casos de investigación de la variabilidad en plantas las 'similitudes' o 'diferencias' entre individuos se describen o se refieren en algún momento a las características morfológicas, ya que esa es la primera percepción natural. Son muy pocos los grupos de vegetales que cuentan con el sustento de la descripción integral del modo de transmisión o heredabilidad y la evolución de las características morfológicas que las distinguen. Se entiende, por tanto, que los caracteres morfológicos son indicadores que pueden portar errores sistemáticos en el análisis de la variabilidad y que es preciso contrarrestar con el cuidado en su selección y análisis (Perrier, 1998).

Los marcadores morfológicos que son distinguibles por observación directa han sido utilizados como marcadores de la diversidad y reflejan el genoma y sus regulaciones biológicas y medioambientales. Estas características son raramente monogénicas y casi siempre resultan de la interacción de varios genes, además, presentan la ventaja esencial de estar disponibles en gran cantidad, no requieren de sofisticados sistemas de medición y son registrados directamente en el campo. La posibilidad de contar con gran número de caracteres y repeticiones ayuda en el tratamiento analítico.

Perrier (1998) señala algunas reglas de selección de características que pueden contribuir a la utilidad de los marcadores morfológicos:

- Retener los caracteres cualitativos a modalidades poco numerosas, exclusivos y distinguibles sin ambigüedad, tratando de aproximarse a marcadores monogénicos o gobernados por unos pocos genes;
- preferir la observación de las partes de órganos no sometidos a fuertes presiones selectivas, naturales o agronómicas, para evitar lo que los taxónomos llaman 'homoplastias' o características ligadas a una respuesta adaptativa idéntica;
- evitar mediciones cuantitativas de altura, longitud, ancho y diámetro, que son fuertemente afectadas por las condiciones del medio; y
- estandarizar las condiciones de observación.

Se debe señalar que los trabajos fundamentales de clasificación han estado basados en las características morfológicas y que en la mayoría de los grupos de plantas

representan la única vía de acceso al estudio de la diversidad natural y agronómica.

Análisis de la información morfológica

La medición y el análisis de la variabilidad morfológica en plantas no es un proceso único y puede tener diferentes finalidades. Algunos de los objetivos más comunes que se pueden abordar cuando se cuenta con información morfológica consisten en describir la estructuración de poblaciones, medir la variabilidad de una colección con fines de clasificación de especies o accesiones, identificar duplicados y detectar combinaciones y características particulares en individuos.

Estructuración de la diversidad de poblaciones

Para evitar confusión, el polimorfismo de marcadores morfológicos se llamará variabilidad genética, porque en genética de poblaciones la diversidad se refiere a la descrita con marcadores moleculares. La variabilidad genética se refiere esencialmente a los componentes de la varianza de caracteres cuantitativos. Los caracteres cualitativos cuentan con parámetros equivalentes dentro del campo de la estadística no-paramétrica, en caso de ser necesaria la estimación de la variabilidad con estos tipos de descriptores (Kremer, 1994).

La diversidad genética o coeficiente de Nei (H) es el parámetro más generalizado en poblaciones y se traduce como la probabilidad que dos genes en una población sean diferentes, así:

$$H = 1 - \sum p_i^2$$

donde,

p_i es la frecuencia media del alelo i en los loci considerados, y H es equivalente a la heterozigosidad teórica bajo los supuestos de Hardy-Weinberg.

Cuando se utilizan los descriptores morfológicos, el nivel de variabilidad de una población puede ser estimado a partir de los diferentes componentes de la varianza genotípica (VA , varianza aditiva; VD , varianza de dominancia; y VG , genotípica). Estos componentes son estimados a partir de las covarianzas con pruebas de descendencias.

Para el caso más simple de un locus con dos alelos, se puede demostrar que hay una relación entre la variabilidad genética y la diversidad en una población, o sea:

$$VA = Ha^2$$

donde, a es el efecto mediano de sustitución de dos alelos;

$$VD = H^2d^2$$

donde, d es el efecto de dominancia generado por dos alelos;

$$VG = H(a^2 + Hd^2)$$

donde, H es la diversidad genética de Nei (1977).

Sin embargo estas igualdades no son generalizables al caso de un locus en varios alelos, a pesar de que los componentes de la diversidad intervienen en la expresión de

las diferentes varianzas. Por el contrario, pueden ser generalizadas a un carácter gobernado por varios loci, cada uno de ellos bialélicos. Esto permite derivar que la variabilidad genética de un carácter fenotípico depende de la diversidad genética ligada con las frecuencias alélicas y del efecto aditivo o dominante de dichos alelos.

Como las varianzas dependen de la unidad de medida, para facilitar las comparaciones a posteriori se recomienda utilizar el coeficiente de variación o la heredabilidad en sentido estricto o amplio, que corresponde a la varianza aditiva o genotípica.

Parámetros de diferenciación entre poblaciones

La estructuración genética de un conjunto de poblaciones es descrita utilizando los estadísticos F de Wright (1965) a nivel alélico. Nei (1977) lo generalizó para un conjunto de loci. El coeficiente de diferenciación genética entre poblaciones (GST) se define de la manera siguiente:

$$GST = 1 - HS / HT$$

donde, HS y HT son las medias (del conjunto de loci) de las diversidades intrapoblacional y total, respectivamente.

La variabilidad fenotípica de un carácter cuantitativo puede ser descompuesta según un modelo jerárquico en un componente interpoblacional (V_{inter}) y otro intrapoblacional (V_{intra}). La diferenciación entre poblaciones puede ser descrita con la relación:

$$t = V_{inter} / (V_{inter} + V_{intra})$$

Para un carácter neutro es posible establecer una relación entre GST y t (Rogers y Harpending, 1983):

$$V_{inter} = 2GSTVA_t$$

$$V_{intra} = (1 - GST)VA_t$$

donde, VA_t es la varianza aditiva de la población en equilibrio Hardy-Weinberg constituida de todas las poblaciones, lo que resulta en:

$$t = 2GST / (1 + GST)$$

En general, y debido a los bajos valores de diferenciación, el valor de t será siempre superior al de GST . Para que ambos estimadores sean comparables, se recomienda tomar t' como el coeficiente de diferenciación para características morfológicas:

$$t' = t / (2 - t)$$

Organización de las accesiones

La descripción de la variabilidad morfológica tiene por objetivo identificar, describir o explicar una eventual organización de individuos. Si estos individuos son simples y

fácilmente distinguibles, la organización se revelará con pocos caracteres. Pero si los individuos son numerosos y complejos, la distinción debe ser el resultado de una caracterización con un gran conjunto de atributos para no correr el riesgo de sesgo o de responder sólo a un interés en particular. Este es el caso frecuente en la mayor parte de los estudios con colecciones de plantas.

En torno a este problema se han desarrollado diversos métodos numéricos. El análisis de este tipo de información se reduce, muy a menudo, al análisis de matrices de individuos x variables. Este análisis tiene como objetivo describir la estructuración de los individuos y las variables a través de sus divergencias o similitudes y sus relaciones. Los análisis multivariados factoriales y de clasificación se utilizan frecuentemente como complementarios (Perrier et al., 1999).

Análisis factoriales

Los métodos multivariados tipo factoriales describen el comportamiento o las tendencias de un conjunto de individuos identificados con una serie de variables o descriptores. Los análisis componentes principales (ACP) y factorial de correspondencias (AFC) (simples y múltiples) pertenecen a este tipo. Independiente del tipo de análisis factorial utilizado, el objetivo principal es generar una representación gráfica de los individuos en un espacio de un número menor de dimensiones que las n variables que se midieron (en general dos o tres). Los individuos, las variables o ambos simultáneamente, son representados en planos de dos o tres dimensiones en función de sus proximidades. La proximidad y el criterio de proyección utilizados por los diferentes métodos factoriales varían en función de la naturaleza de las variables (cuantitativas o cualitativas). El fundamento y ventaja de estos análisis consiste en que se parte de la definición de proximidades entre individuos y variables, sin tomar en cuenta a priori su pertenencia a grupos definidos o poblaciones. La distancia utilizada en el ACP es la euclidiana clásica. La selección del número de ejes de proyección se realiza tratando de minimizar la pérdida de información, pero captando la mayor varianza total posible. La distancia en el AFC es ji-cuadrada (χ^2) y en él se busca, de forma análoga, maximizar la variabilidad total descrita con pocos ejes o factores (Segura, 2001).

Análisis de clasificación

La clasificación en el lenguaje corriente se refiere a dos posibles acciones. La primera, al hecho de agrupar los individuos en conjuntos homogéneos, uno de los objetivos de los grandes taxónomos. La segunda, al ordenamiento lógico de los individuos, un objetivo que aborda las relaciones de filiación entre entidades filogenéticas.

Ambas aproximaciones pueden generar diagramas ramificados similares, aunque su información será diferente. En el primer caso, los nudos corresponderán a conceptos de agrupamiento, mientras que en el segundo se suponen unidades parentales no observadas aunque supuestas.

El problema de la divergencia genética retoma parte de ambas aproximaciones porque trata de agrupar a los individuos similares y de relacionarlos de acuerdo con la filiación.

Para el análisis de la organización de la variabilidad genética en plantas se distinguen dos enfoques: el fenético que se refiere a la descripción de la organización a partir de medidas objetivas de divergencias estimadas de un conjunto de observaciones, y el cladístico que intenta identificar y utilizar caracteres informativos que han sido

heredados y que permiten describir las rutas de evolución. La naturaleza de las observaciones morfológicas y el grado de conocimiento evolutivo de la mayor parte de grupos de plantas no permite seguir la aproximación cladística. Los métodos que se utilizan generalmente en el estudio de divergencias entre individuos siguen una aproximación fenética, pero no son citados como tales en los trabajos científicos escritos porque se supone que son conocidos (Perrier, et al., 1999).

La etapa inicial de los análisis de clasificación corresponde a la definición de la medida de divergencia entre individuos. La selección de esta medida debe corresponder a la naturaleza de los datos y al origen de los individuos. Las distancias euclidiana, euclidiana media y 'city-block' se aplican muy a menudo a datos cuantitativos; las distancias de Rogers y Tanimoto, Sokal y Michener y Sokal y Sneath (No. 1) a datos cualitativos de modalidades; la distancia Ji-cuadrada a datos cualitativos ordinales. Las distancias de Dice, Jaccard o de Sokal y Sneath (No. 2) se usan cuando los datos son binomiales, como es el caso muy frecuente, mas no único, de los descriptores moleculares.

Para la representación arbórea ('clusters') de los individuos, una vez calculadas las distancias, se debe elegir un método de agrupamiento. Estos métodos se refieren a los procesos iterativos que aglomeran a los individuos y que definen a los vecinos en las ramas representadas. Criterios como el de la distancia más próxima, la distancia media (UPGMA) o la media ponderada (WPGMA) son algunos ejemplos. Un método muy utilizado actualmente, pero que desafortunadamente pocos paquetes informáticos de análisis lo incluyen, es el de Neighbor-joining, propuesto por Saitou y Nei (1987), que utiliza un criterio de media no-ponderada y que va juntando individuos y recalculando las nuevas clasificaciones sin los elementos ya tratados. Las representaciones 'desracinadas' de este algoritmo describen de manera más clara las divergencias entre individuos cuando no se supone, porque no es necesario, un ancestro común.

Referencias

- Jørgensen, P. M.; Lawesson, J. E., y Holm-Nielsen, L. B. 1984. A guide to collecting passion flowers. *Ann. Missouri Bot. Gard.* 71:1172-1174.
- Nei, M. 1977. F statistics and analysis of gene diversity in subdivided populations *Ann. Hum. Genet.* 41: 225-233.
- Perrier, X. 1998. Analyse de la diversité génétique: Mesures de dissimilarité et représentations arborées. Doctorat, Université Montpellier II. Montpellier. 192 p.
- Saitou, N. y Nei, M. 1987. Neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406-425.
- Wright S. 1965. The interpretation of populations structure by F-statistics with special regard to systems of mating. *Evolution* 19:395-420.
- Tillett, S. S. 1988. *Passionis passifloris* II. En: Ernstia, (ed.). Terminología. Herb. Fac. Agron. Univ. Central de Venezuela, Maracay. p. 1-49.
-

Análisis multivariado en estudios de variabilidad genética

*Wilfredo Rojas**

La información generada en la caracterización y evaluación de las colecciones de germoplasma proporciona una idea del tipo de material genético que se conserva en el banco y su utilización es más importante cuando se estudia su variabilidad.

La diversidad y la variabilidad genéticas son términos alternativos para representar la variación genética. Se sugiere que diversidad sea utilizada para indicar la sumatoria de la información genética potencial conocida y desconocida, y la variabilidad para indicar una porción de la diversidad capturada o disponible (Vilela-Morales y Candeira, 1996). Por tanto, el término variabilidad genética es el más indicado para el estudio de colecciones de germoplasma.

Taba (1991) considera que la variabilidad genética de un cultivo puede ser estudiada por las técnicas moleculares más avanzadas o a través de rasgos morfológicos; la diferencia entre ellas consiste en que la primera no considera la interacción genotipo x ambiente, sin embargo, ambas son complementarias. Fairbanks et al. (1992), por otra parte, califican la influencia del medio ambiente como una desventaja en las caracterizaciones morfológicas, ya que no permite representar verdaderas similitudes o diferencias genéticas.

Souza y Sorrells (1991) sugieren que la variabilidad y las relaciones genéticas entre una población en particular pueden ser medidas por la similitud de algún número de caracteres cuantitativos considerados como parámetros agronómicos de las plantas, lo cual asume que las diferencias entre los caracteres de los genotipos reflejan la divergencia genética existente en ellos.

Con la información que se genera de la caracterización y evaluación de las colecciones de germoplasma es posible medir su variabilidad genética o conocer qué tan variables son las accesiones que lo conforman, a través de la similitud o diferenciación de los rasgos que caracterizan a cada una de ellas. Así mismo, cuando se agrupan accesiones de un germoplasma de acuerdo con su similitud morfológica o fenotípica, es de suponer que gran parte de ésta es genética.

Las técnicas estadísticas multivariadas son herramientas muy útiles para caracterizar germoplasma, debido a que básicamente permiten describir o agrupar un conjunto de accesiones, tomando en cuenta simultáneamente varias características, sin dejar de considerar la relación existente entre todos los caracteres en estudio.

Referencias

- Fairbanks, D.; Robinson, L.; Burgener, K.; Andersen, W.; López, D.; Ballon, E.; Bernal, R.; Peña, R.; Sánchez, J.; y Pimentel, A. 1992. Caracterización de material genético disponible de quinua. En: Morales, D. y Vacher, J. (eds.). Actas del VII Congreso Internacional sobre Cultivos Andinos. La Paz, Bolivia. 4-8 de julio de 1991. p. 19-21.

* Especialista en Recursos Fitogenéticos, Fundación PROINPA, Regional Altiplano, La Paz-Bolivia.

- Souza, E. y Sorrells, M. E. 1991. Relationships among 70 North American oat germplasms: I. Cluster analysis using quantitative characters. *Crop Sci.* 31:599-605.
- Taba, S. 1991. Caracterización y evaluación de germoplasma de maíz. En: Castillo, R.; Estrella, J; y Tapia, C. (eds.). *Técnicas para el manejo y uso de recursos genéticos vegetales*. Departamento de Recursos Fitogenéticos, Instituto Nacional de Investigaciones Agropecuarias, INIAP. Quito, Ecuador. p. 161-166.
- Vilela-Morales, E. A. y Candeira, A. C. 1996. Principios genéticos para recursos genéticos. En: Puignau, J. P. (ed.). *Curso sobre conservación de germoplasma vegetal*. Brasilia, Brasil. 19 - 30 de septiembre de 1994. Montevideo: Diálogo XLV - IICA-Procisur. p. 35-48.
-

Bibliografía Consultada

- Argüelles, J. 1990. Selección de variables en cebada mediante el análisis de componentes principales. *Revista ICA* 25(4):355-365.
- Anderberg, M.R. 1973. *Cluster analysis for applications*. Nueva York, Academic Press. 361 p.
- Cochran, W. 1954. Some methods for strengthening the common χ^2 tests. *Biometrics* 10:417-451
- Crisci, J.V. 1983. Introducción a la teoría práctica de la taxonomía numérica. Secretaría de la Organización de Estados Americanos (OEA). Washington, D.C. p. 39-67.
- Chalfield, C. y Collins, A. J. 1986. *Introduction to multivariate analysis*. 3th. edition. Chapman and Hall. Nueva York. 246 p.
- De la Sota, E. R. 1973. La taxonomía y la revolución de las ciencias biológicas. Secretaría de la Organización de Estados Americanos (OEA). Washington, D.C. 86 p.
- Debouck, D.G. e Hidalgo, R. 1985. Morfología de la planta de frijol. En: López, M.; Fernández, F; y Van Schoonhoven, A. (eds.). *Fríjol investigación y producción*. CIAT, Cali, Colombia. p. 7-41.
- FAO, 1996. Informe sobre el estado de los recursos fitogenéticos en el mundo. En: Conferencia Técnica Internacional sobre los Recursos Fitogenéticos. Leipzig, Alemania, 17-23 junio de 1996. Roma, Italia.
- FAO, 1997. Conservación y utilización sostenible de los recursos fitogenéticos para la alimentación y la agricultura. Plan de Acción Mundial. Roma, Italia. 10 p.
- Fajardo, D.; Angel, F.; Grum, M.; Tohme, J.; Lobo, M; Roca, W; y Sánchez, I. 1998. Genetic variation analysis of the genus *Passiflora* L. using RAPD markers. *Euphytica* 101 (3): 341-347.
- Fienberg, S. 1977. *The analysis of cross-classified data*. Cambridge, MA. MIT Press.
- Franco, J; Crossa, J.; Villaseñor, J.; Taba, S; y Eberhart, S. A. 1998. Classifying genetic resources by categorical and continuous variables. *Crop Sci.* 38:1688-1696.
- Franco, J; Crossa, J.; Diaz, J.; Taba, S.; Villaseñor, J.; y Eberhart, S. A. 1997a. A sequential clustering strategy for classifying gene bank accessions. *Crop Sci.* 37:1656-1662.
- Franco, J; Crossa, J.; Villaseñor, J.; Taba, S.; y Eberhart, S. A. 1997b. Classifying Mexican maize accessions using hierarchical and density search methods. *Crop Sci.* 37:972-980.
- Frankel, O. H. 1970. Evaluation and utilization – Introductory remarks. En: Frankel, O.H. y Bennet, E. (eds.). *Genetic resources in plants – Their exploration and conservation*. International Biological Programme. Blackwell Scientific Publications, Oxford and Edinburg. IBP Handbook no. 11. p. 395-401.
- Gower, J. C. 1967. Multivariate analysis and multivariate geometry. *The Statistician* 17:13-28.
- Gower, J. C. 1967. A comparison of some methods of cluster analysis. *Biometrics* 23:623-637.
- Gower, J.C.. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27:857-874.
- Harlan, J. R. 1975. Prologue: The golden age. En: *Crops and man*. Amer. Soci. Agron., Crop Sci. Soci. Amer. p.1-32.
- Harch, B. D. 1995. *Estatistical evaluation of germplasm collections*. PhD. Thesis. University of Queensland, Department of Agriculture. 180 p.
- Hidalgo, R. 1991. CIAT's World *Phaseolus* collection. En: Van Schoonhoven, A. y Voysest, O. (eds.). *Common beans: Research for crop improvement*. CIAT, Cali, Colombia. p. 163-178.
-

- Hoyts, E. 1992. La clasificación de las plantas: Taxonomía o sistemática. En: Conservando los parientes silvestres de las plantas cultivadas. Editorial Addison-Wesley, Iberoamericana, S.A., Delaware, E.U. 52 p.
- Jaramillo, S. y Baena, M. 2000. Material de apoyo a la capacitación en conservación ex situ de recursos fitogenéticos. Instituto Internacional de Recursos Fitogenéticos (IPGRI). Cali, Colombia. 122 p.
- Johnson, R. A. y Wichern, D. 1988. Applied multivariate statistical analysis. 2nd. Edition. Prentice Hall, Englewood, Nueva Jersey, E.U. 607 p.
- Kendall, M. y Stuart, A. 1979. The advanced theory of statistics. Nueva York. Macmillan Publ. Co. Inc. Volumen 2. s.p.
- Krarup, H. A. 1984. Organización de la variabilidad genética en poblaciones de plantas. En: Anales. Simposio sobre recursos Fitogenéticos. Universidad Austral de Chile e IBPGR. Valdivia. Chile. p. 20-27.
- Kremer, A. 1994. Diversité génétique et variabilité des caractères phénotypiques chez les arbres forestiers. Genet. Sel. Evol. 26:1:105s-123s.
- López, A. J. 1991. Descripción sistemática y parámetros genéticos para características cualitativas y cuantitativas en la colección de batata *Ipomoea batata* (L.) Lam. del CATIE. Tesis Mag. Sc. Turrialba, Costa Rica. 128 p.
- Manly, B. F. 1994. Multivariate statistical methods. 2nd. edition. Chapman and Hall. Nueva Zelandia. 215 p.
- Márquez, J. M. 1992. Caracterización sistemática, parámetros genéticos e índices de selección de la colección de jícama (*Pachyrhizus erosus* L. Urban) del CATIE. Tesis Mag. Sc., CATIE, Turrialba, Costa Rica. 142 p.
- Martínez, W. O. 1995. Análisis estadístico en biología molecular: Uso y aplicación en poblaciones vegetales. En: Memorias Simposio Internacional de Estadística en Agricultura y Medio Ambiente. CIAT, Cali, Colombia. p. 152-170.
- Onyilagha, J. C. 1986. Numerical analysis of variation among *Nigerian Dioscorea* rotundata accessions. Euphytica 35:413-419.
- Peeters, J. P. y Martinelli, J. A. 1989. Hierarchical cluster analysis as a tool to manage variation in germoplasm collections. Theoret. Appl. Genet. 78:42-48.
- Perrier, X. 1998. Analyse de la diversité génétique: Mesures de dissimilarité et représentations arborées. Thèse doctorat. Université de Montpellier II. Montpellier. 192 p.
- Pielou, E.C. 1984. The interpretation of ecological data. A primer on classification and cordination. John Wiley & Sons. Nueva York. p. 136-151.
- Podani, J. 1993. Syntax. Computer programs for multivariate data analysis in ecology systematic. Version 5.0. Budapest, Hungria. Scient. Publi. 104 p.
- Querol, D. 1987. Genetic resources. A practical guide to their conservation. Zed Books Ltd. Londres y Nueva Yersey. Penang, Malasia. p. 173-185
- Rohlf, F. J. 1998. NTSYS. Numerical taxonomy and multivariate analysis system. Version 2.0, Exeter Publ., Setauket, Nueva York. 31 p.
- Sánchez, I; Angel, F; Grum, M; Duque, M. C.; Lobo, M.; y Tohme, J. 1999. Variability of chloroplast DNA in the genus *Passiflora* L. Euphytica 106 (1):15-26.
- Segura, S. D. 2001. Distribution et organisation de la diversité des Passiflores Andines (sous-genre *Tacsonia*). Tesis de doctorado. ENSA. Rennes. p. 135.
- Scholotzhauer, S. D. y Littell, R. C. 1987. SAS System for elementary statistics analysis. SAS Institute, North Carolina, E.U. 417 p.
-

- Singh, S. P. 1985. Conceptos básicos para el mejoramiento del frijol por hibridación. En: López; M.; Fernández, F.; y Van Schoonhoven (eds.). Frijol investigación y producción. CIAT, Cali, Colombia. p. 109-126.
- Sokal, R. R. y Sneath, H. A. 1963. Numerical taxonomy. The principles and practice of numerical classification. Freeman and company . San Francisco, E.U. 358 p.
- Sober, E.; 1993. Philosophy of biology. Bourder Westwicw Press. p. 143-183.
- Tohme, J.; Jones, P.; Beebe, S.; Iwanaga, M.; y Toro, O. 1993. Forming a core collection of *Phaseolus vulgaris* L. En: *Phaseolus* beans advanced biotechnology research network. CIAT, Cali, Colombia. p. 118-122.
- Tapia, C. G. 1998. Caracterización morfológica y molecular de la diversidad genética de la colección de *Pachyrhizus tuberosus* (LAM.) Spreng. del CATIE. Tesis Mag. Sci., CATIE. Turrialba, Costa Rica. 157 p.
- Ward, J. H., Jr. 1963. Hierarchical grouping to optimize an objective function. J. Amer. Statist. Assoc. 58:236-244.
-